UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE
RIBEIRÃO PRETO
DEPARTAMENTO DE CONTABILIDADE
PROGRAMA DE PÓS-GRADUAÇÃO EM CONTROLADORIA E CONTABILIDADE

LUIZA DE MIRANDA CABRAL

Movimento ocular e desafios cognitivos na compreensão de Relatórios Populares

ORIENTADOR: PROF. DR. ANDRÉ CARLOS BUSANELLI DE AQUINO

Ribeirão Preto

2019

Prof. Dr. Vahan Agopyan
Reitor da Universidade de São Paulo

Prof. Dr. André Lucirton Costa
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Fabiano Guasti Lima
Chefe do Departamento de Contabilidade

Prof. Dr. Amaury José Rezende
Coordenador do Programa de Pós-Graduação em Controladoria e Contabilidade

LUIZA DE MIRANDA CABRAL

Movimento ocular e desafios cognitivos na compreensão de Relatórios Populares

Dissertação apresentada ao Programa de Pós-Graduação em Controladoria e Contabilidade da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências. Versão Corrigida. A original encontra-se disponível no Serviço de Pós-Graduação da FEA-RP/USP.

ORIENTADOR: PROF. DR. ANDRÉ CARLOS
BUSANELLI DE AQUINO

Ribeirão Preto

2019

Versão Corrigida. A original encontra-se disponível na FEA-RP/USP

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

Cabral, Luiza de Miranda

Movimento ocular e desafios cognitivos na compreensão de Relatórios Populares / Luiza de Miranda Cabral. - Ribeirão Preto, 2019.

87 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2019.

Orientador: André Carlos Busanelli de Aquino

1. Contabilidade governamental. 2. Contabilidade Pública. 3. Comunicação da contabilidade. 4. Controladoria. I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

FOLHA DE APROVAÇÃO

CABRAL, L. M. (2019). **Movimento ocular e desafios cognitivos na compreensão de Relatórios Populares**. Dissertação (Mestrado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2019.

Aprovada em: ____ / ____ / 2019

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dra. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Dedico este trabalho às minhas filhas,
Ana Rosa e Cecília, com todo meu coração.

Que um dia vocês possam realizar seus
próprios sonhos, sabendo que a vida é um
eterno aprender e descobrir.

AGRADECIMENTOS

Agradeço a Deus, porque sei que ele está no controle de todas as coisas, e que me deu essa oportunidade. Obrigada, Senhor, por essa benção.

A minha família, pelo apoio incondicional. Em especial ao meu marido, meus pais, meu sogro e minha sogra. Nada disso seria possível sem sua ajuda e carinho.

As minhas bebês, Ana Rosa e Cecília, pelos doces sorrisos. Eu sou por vocês.

Ao Prof. Dr. André Aquino, que sempre me ensinou muito mais do que teorias e métodos. Obrigada pelo exemplo de pessoa, profissional e amigo. E por acreditar em mim quando pensei que não seria capaz. Foi um privilégio trabalhar ao seu lado.

Ao Prof. Dr. Jorge Henrique Caldeira de Oliveira pelos ensinamentos e apoio com o equipamento de registro do movimento ocular (*Tobii Eye Tracker* modelo X1L).

A Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, minha segunda casa, pela oportunidade de realização do curso.

A CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela concessão da bolsa Demanda Social durante o período de realização deste mestrado.

RESUMO

CABRAL, L. M. (2019). **Movimento ocular e desafios cognitivos na compreensão de Relatórios Populares**. Dissertação (Mestrado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2019.

A pesquisa traz evidências de como o movimento ocular de jovens indivíduos está associada à compreensão que eles alcançam em experimentos de leitura de relatórios financeiros de governos. A simplificação é um dos pilares dos relatórios destinados à população de não especialistas, o chamado ‘relatório popular’ ou ‘relatório cidadão’, cuja adoção é defendida internacionalmente. Foi realizado um experimento *between subjects* manipulando a legibilidade de textos e tabelas, seguidos de perguntas de múltipla escolha para medir a compreensão obtida pelo respondente. O experimento utilizou um equipamento de rastreamento ocular para acompanhar a atenção visual dos respondentes em cada região espacial dos estratos de relatório apresentados. Os resultados apontam que a legibilidade melhora o desempenho de tarefas mais simples, associadas a recuperação de informações lidas no texto, mesmo quando o padrão de movimento ocular não é alterado. Porém, parece não ajudar o indivíduo a realizar tarefas cognitivas mais complexas como a interpretação de texto e a extração de informações implícitas. A presença de jargões pode não oferecer dificuldade de leitura local, sobre os mesmos jargões, mas afeta o padrão do movimento ocular sobre o texto como um todo, o que indica uma maior dificuldade no processamento do texto. Quando os indivíduos estão olhando para tabelas, a presença dos jargões faz com que estes olhem mais para as palavras do que para os números. Ou seja, o indivíduo procura entender os termos e não consegue interpretar o que os valores monetários representariam. O principal desafio para a compreensão dos relatórios é a presença de termos técnicos da contabilidade.

Palavras-chave: Atenção visual. Legibilidade. *Eye-tracking*. Relatórios Populares.

ABSTRACT

CABRAL, L. M. (2019). **Eye movements and cognitive challenges in reading comprehension of Popular Reports**. Dissertation (Master's degree on Accounting and Controllershship) – Faculty of Economics, Business Administration and Accounting, University of São Paulo, Ribeirão Preto, 2019.

The research provides evidence on how young individuals' eye movements are associated with their understanding of government financial reports. Simplification is one of the main characteristics of the reports dedicated to the non-specialist population, the so-called 'popular report' or 'citizen report', as advocated internationally. An experiment between subjects was performed manipulating the readability of texts and tables, followed by multiple choice questions designed to measure comprehension of the passages. The experiment used an eye tracking device to track the respondents' visual attention in each spatial region of the presented reporting stratus. Results show that readability improves the performance of simpler tasks such as recalling the information they've just read, even when eye movement pattern does not change. However, lower readability didn't seem to help individuals to perform more complex cognitive tasks, like text interpretation and retention of implicit information. Besides, the presence of jargon may not cause local reading difficulty but it does affect the eye movements over the text as a whole, which is a clear sign of greater difficulty in processing the information. Finally, when individuals are looking at tables, the presence of jargons causes individuals to look more at words than at numbers. Therefore, individual tries to understand the terms instead of focusing on the monetary values presented in the tables. The research shows that the main challenge for understanding public financial reports is the presence of technical accounting terms.

Keywords: Eye movements. Eye-tracking. Popular Reporting.

SUMÁRIO

1. INTRODUÇÃO	12
2. LITERATURA RELEVANTE.....	15
2.1 O Relatório Popular	18
3. QUADRO TEÓRICO E DESENVOLVIMENTO DAS HIPÓTESES.....	22
3.1 Legibilidade e Compreensão.....	22
3.2 O Movimento Ocular durante a Leitura	25
3.3 Hipóteses de Pesquisa.....	27
4. METODOLOGIA.....	28
4.1 O desenho do experimento.....	29
4.2 O Equipamento de rastreamento ocular (<i>Eye-Tracker</i>).....	30
4.3 As métricas geradas pelo <i>Eye-tracker</i>	31
4.4 Jovens Adultos e o “Cidadão comum”	31
4.5 Amostragem	33
4.6 Identificação e validação dos jargões na área (termos técnicos).	35
4.7 O Desenho dos Testes 1, 2 e 3	36
4.7.1 O Teste 1 – Legibilidade sintática e léxica dos textos (relatos de desempenho de governos)	37
4.7.2 O Teste 2 – Jargões nos textos (relatos de desempenho de governos).....	39
4.7.3 O Teste 3 – Jargões em tabelas (demonstrações contábeis de governos).....	40
4.8 Aplicação do experimento.....	41
5. RESULTADOS E ANÁLISES	43
5.1 Resultados e Análises do Teste 1: Legibilidade sintática e léxica.....	43
5.2 Resultados e Análises do Teste 2: Legibilidade pela presença de jargões	46
5.3 Resultados e Análises do Teste 3: Legibilidade de tabelas com jargões	51
6. IMPLICAÇÕES	54
7. CONCLUSÕES.....	57
REFERÊNCIAS	60
APÊNDICE A – O protocolo do experimento	64
APÊNDICE C – Perguntas atenção e compreensão testes 1 e 2	68
APÊNDICE D – Mapas de calor dos testes T1 a T3	69
APÊNDICE E – Tabelas e mais dados com análises descritivas e estatísticas	72
APÊNDICE F – Dados sobre a validação da dificuldade dos jargões contábeis	85
ANEXO 1 – Ofício de Aprovação do Projeto de Pesquisa pelo CEP	87

1. INTRODUÇÃO

A importância da transparência e da *accountability* no setor público já foi reconhecida em inúmeros estudos, sob diferentes perspectivas (e.g. Hood, 2010; Lindstedt e Naurin, 2010; Kolstad e Wiig, 2009; Heald, 2003; Justice e Dulger, 2009). De acordo com o IPSASB (2013) (*International Public Sector Accounting Standards Board* – Comitê internacional que elabora normas de contabilidade pública), fornecer informações relevantes para os usuários e permitir a *accountability* são os principais objetivos da contabilidade pública. Várias pesquisas discorrem sobre a importância do uso e da utilidade dessas informações para os cidadãos (e.g. Brusca e Montesinos, 2006; Rodríguez Bolívar e Galera, 2016; AGA, 2012).

Apesar de se saber que os governos deveriam fornecer informações inteligíveis à sociedade, pessoas que não tem conhecimentos em contabilidade pública consideram os relatórios tradicionais ilegíveis, muito extensos e/ou desinteressantes (Jordan, Yusuf, Berman & Gilchrist, 2017). Por isso, o Conselho de Normas de Contabilidade Pública (*Government Accounting Standards Board* - GASB) e a Associação de Oficiais de Finanças Públicas (*Government Finance Officers Association* - GFOA) vêm reforçando há mais de 20 anos a necessidade da elaboração de informações especificamente desenhadas para atender as necessidades dos cidadãos (Lee, 2006). Foi buscando endereçar esse problema que um novo modelo de relatório foi idealizado.

O Relatório Popular (tradução livre de *Popular Report*) foi apresentado como uma nova forma de comunicar o desempenho financeiro de governos. Considerado uma importante ferramenta na melhoria da transparência pública, esse tipo de relatório temo objetivo de aumentar *accountability* (Barbera, Borgonovi & Steccolini, 2016). O Relatório Popular é feito para ser facilmente compreendido pelo cidadão que não tem conhecimentos aprofundado ou experiência em finanças públicas. Essencialmente, o RP (Relatório Popular) deve usar uma linguagem mais simples e concisa, além de gráficos, figuras e outros recursos para facilitar o entendimento das informações e promover o engajamento (Yusuf e Jordan, 2012). Não obstante, evidências empíricas que sustentem as suposições assumidas pela literatura de Relatórios Populares são escassas. Estudos realizados com base em grupos focais e questionários (pesquisas de autopreenchimento) (e.g. Yusuf e Jordan, 2012; Jordan, *et al.*, 2016; Paolo, Silvana & Valerio, 2016). Não se sabe, até o momento, de pesquisas que tenham utilizado métodos mais robustos e menos susceptíveis à algum tipo de viés por reflexividade ou de estratégias conscientes.

Assim, a presente pesquisa se propõe a testar se a linguagem mais simples dos relatórios populares poderia de fato aumentar a compreensão do desempenho financeiro de governos entre os não especialistas. Para entender melhor o processo de interação do leitor (de um relatório) com a informação de desempenho financeiro, foi realizado um experimento (*between subjects*) em que jovens adultos realizaram três tarefas de leitura e compreensão de estratos de relatórios de desempenho de governos. Em cada tarefa do experimento observou-se a atenção visual do respondente com uso de um equipamento de rastreamento ocular (*eye tracker*). Este equipamento gera medidas que permitem um melhor entendimento das dificuldades de leitura, pois identifica de forma precisa qual conteúdo oferece maior dificuldade ao ser cognitivamente processado (Rayner, 2006). Para cada uma das três tarefas foi observado o efeito da manipulação da simplificação do estrato do relatório pelo *eye-tracker* que registrou continuamente o movimento ocular dos participantes ao ler os conteúdos dos testes.

As manipulações foram feitas alterando características físicas do texto. No primeiro teste apresenta-se um texto, sendo uma versão com frases e palavras mais curtas. O segundo apresenta outro texto, uma versão com jargões típicos dos relatórios contábeis e a outra com termos menos técnicos. No terceiro apresenta-se uma tabela com dados financeiros em sua versão tradicional (como é normalmente publicado nos relatórios públicos) e uma versão simplificada, em que as palavras mais técnicas (jargões) foram substituídas.

Observando a interação de jovens indivíduos com informações sobre contabilidade e finanças públicas é possível inferir sobre como (e se) o Relatório Popular funciona de fato. As evidências suportam o entendimento de como os textos e tabelas são cognitivamente processados, e se a simplificação das informações ajuda as tarefas cognitivas desempenhadas ao ler os relatórios. O estudo contribui diretamente para a literatura de transparência e *accountability* pois a partir desse entendimento é possível melhorar as diretrizes para elaboração de relatórios mais compreensíveis. Entende-se que esse é um dos principais caminhos para o aumento da transparência efetiva. Contrapondo a defesa da divulgação desenfreada de informações, não necessariamente inteligíveis, e que podem gerar uma transparência ilusória.

Na próxima seção será apresentada a literatura relevante sobre transparência e *accountability* no setor público, contextualizando a idealização e o atual status do Relatório Popular. Em seguida, coloca-se o embasamento teórico e o desenvolvimento das hipóteses. A seção seguinte explica a metodologia usada na pesquisa, incluindo mais detalhes sobre o equipamento de rastreamento ocular, o material apresentado aos respondentes e as tarefas realizadas, além do protocolo de pesquisa. Na quinta seção apresenta-se a análise dos

resultados. Por último, as implicações do estudo serão discutidas, assim como as conclusões e sugestões de pesquisas futuras.

2. LITERATURA RELEVANTE

Pesquisas ressaltam a importância da prestação de contas àqueles que, em última instância, “pagam as contas” dos governos. Quando a gestão pública atua de forma “aberta e responsável”, as decisões políticas podem afetar a vida das pessoas da maneira como elas gostariam (Scholte, 2008). Essencialmente, transparência e accountability levam ao fortalecimento da democracia nos Estados democráticos (Scholte, 2008). São fundamentais para uma boa governança pública (Hood, 2010).

É importante destacar que a definição dos conceitos sobre transparência pública, e o quanto ela beneficia a performance dos governos depende de como o Estado e os processos políticos são vistos (Heald, 2003, p. 753). Em uma visão menos otimista dos efeitos, depois de atingir certo nível, um aumento do nível de transparência leva a diminuição da eficiência dos governos – nessa visão, “muita luz” atrapalha o andamento dos processos executados pela administração pública (Heald, 2003, p.727). Transparência pode ser definida como a abertura e a divulgação geral de todas as informações relevantes (Hood, 2010). No setor público, para que haja ‘transparência efetiva’ o orçamento e as outras informações financeiras precisam ser apresentadas de forma inteligível, útil e acessível aos cidadãos comuns (Justice e Dulger, 2009). Em outras palavras, a transparência efetiva se refere a divulgação de conteúdos que servem como instrumento no processo de avaliação das instituições públicas e governantes (Ceneviva & Farah, 2012).

É possível colocar os conceitos de transparência e accountability como parte essencial de um modelo de vigilância, no qual os cidadãos são responsáveis pelo monitoramento das autoridades (Eckersley, Ferry e Zakaria, 2014). Nesse sistema, supõe-se que muitas pessoas observam o comportamento de outras poucas (Mathiesen, 1997). Com base na interpretação do desempenho dos governantes, essa massa de pessoas deveria punir ou recompensar, principalmente por meio do voto. Entretanto, devido à natureza e ao volume dos dados, os cidadãos não são capazes e/ou não estão dispostos a desempenhar sua função (Eckersley et al., 2014). Assim, se as informações ininteligíveis para os cidadãos, mesmo que os governos ressoem uma retórica de políticas que aumentam a transparência, o modelo proposto pode simplesmente não funcionar.

Por isso um aumento na quantidade de dados não necessariamente leva a uma maior transparência efetiva. A chamada ‘transparência ilusória’ é o resultado da diferença entre

‘transparência nominal’ e ‘transparência efetiva’ (Heald, 2003). Ou seja, mesmo que mais e mais dados sejam divulgados (transparência nominal), se o conhecimento externo que se obtém sobre como as finanças públicas estão operando (transparência efetiva) não aumenta, o que se está produzindo é uma ‘ilusão de transparência’ (Heald, 2003).

Em outras palavras o problema da transparência ilusória se dá quando os dados divulgados não favorecem uma interpretação correta e de qualidade do desempenho dos governantes. Ocorre principalmente de duas formas: diminuindo a comparabilidade dos dados ou levando a interpretações tendenciosas. Entre outros, a não divulgação de análises relevantes, a mudança das nomenclaturas e definições dos termos utilizados, a divulgação de dados incompletos e a alocação de eventos importantes ou dívidas “off-balance sheet” são exemplos de práticas que prejudicam as avaliações tão necessárias¹ (Heald, 2003).

Pode até haver a sensação de que um governo está sendo transparente ao divulgar muitas informações. Mas, se o material disponibilizado não facilita os processos de avaliação e responsabilização, não haverá accountability (Ceneviva & Farah, 2012). Assim, observa-se que os conceitos – transparência e *accountability* - estão intimamente ligados. *Accountability* pode ser definido como sinônimo de responsabilidade e responsividade (Campos, 1990). É efetivo quando os indivíduos e entidades não apenas explicam o que vem sendo feito com os recursos públicos, mas quando também se fazem responsáveis por sua gestão e são passíveis de punição pelas más decisões (Campos, 1990).

Para realizar o controle social dos gastos públicos a sociedade precisa de ferramentas especiais, o que, por sua vez, depende de uma supervisão organizada das contas públicas por parte da sociedade (Campos, 1990). Um exemplo desse tipo de supervisão organizada são os observatórios sociais, que atuam no monitoramento das contas públicas e na educação financeira, promovendo uma cultura de cidadania (Observatório Social do Brasil, 2018)

Se o *accountability* parte do alto escalão para pessoas e comunidades - “de cima para baixo” -, *accountability* e transparências poderiam ser vistos como gêmeos siameses (Hood, 2010). Porém, para estabelecer uma relação entre os dois termos é preciso uma análise

¹ No Reino Unido, por exemplo, a divergência entre retórica e prática ocorre, dentre outras formas, pela não divulgação de análises relevantes e alocação de dívidas “off-balance sheet” (Heald, 2003).

filosófica, perpassando o debate sobre o que é *accountability* efetivo, formal, nominal e periférico (Hood, 2010).

Nas últimas décadas, de uma forma geral, as reformas administrativas passaram a considerar a necessidade de apresentar informações mais transparentes e úteis para a sociedade (Rodríguez Bolívar & Galera, 2016). Um marco desse movimento é a Normativa 34 (*Statement 34*), emitida nos Estados Unidos, em 1999, pelo o Comitê de Normas Contábeis do Setor Público (*Government Accounting Standards Board – GASB*) (Rodríguez *et al.*, 2016). O texto definiu um novo modelo para relatórios contábeis de estados e municípios e uma das mudanças foi a inclusão do cidadão como um usuário da informação contábil (Dean, 2002). Outra alteração importante foi que passou a ser obrigatório a emissão do Relatório de Gestão e Análise (tradução livre do termo em inglês *Management Discussion and Analysis - MD&A*) (Marsh & Montondon, 2005a e 2005b). As medidas tinham o objetivo de promover mais transparência e *accountability*. Entretanto, apesar das mudanças e de um forte discurso político a favor do *accountability* voltado ao cidadão, as organizações públicas invariavelmente ainda divulgam relatórios opacos, tendenciosos, incompletos, inacessíveis ou parcialmente acessíveis, excessivos ou irrelevantes (Kolstad & Wiig, 2009).

Estudos na área de administração pública vêm tentando identificar e entender o uso dessas informações pelos cidadãos. Quando se trata do comportamento dos eleitores ao votar, há evidências de que os relatórios financeiros divulgados influenciam os resultados das eleições, especialmente os dados da gestão dos gastos públicos (Brusca & Montesinos, 2006). Porém, há outros estudos que evidenciam um comportamento contrário. Nos EUA, por exemplo, a maioria dos cidadãos pensam que o Relatório Financeiro Anual Completo (principal relatório emitido pelos entes governamentais) é muito extenso e incompreensível (Sharp, Carpenter e Sharp, 1998; AGA, 2012). Essas pesquisas explicam que a complexidade dos relatórios os torna inacessíveis, não servindo como informação para tomada de decisão em relação ao voto, por exemplo.

Os relatórios financeiros tradicionais não são conhecidos por serem fáceis de usar. Eles são longos, complexos e cheios de jargões técnicos, que são ininteligíveis para o cidadão comum (Jordan *et al.*, 2017, p. 1, tradução livre).

De fato, os relatórios financeiros tradicionais de organizações públicas têm sido amplamente criticados por não serem úteis para os cidadãos em geral. A maioria dos relatórios

não é elaborada considerando indivíduos sem experiência em contabilidade (Yusuf *et al.*, 2013). Sendo assim, esses instrumentos não são capazes de promover o controle social, não facilitam o monitoramento do desempenho por parte dos cidadãos e, portanto, não viabiliza *accountability* vertical efetivo (Carpenter e Sharp, 1992; Lee, 2006; Yusuf *et al.*, 2013).

Pesquisas nos Estados Unidos (Sharp, *et al.*, 1998; AGA, 2012), Inglaterra (Eckersley *et al.*, 2014) e Alemanha (Haustein e Lorson, 2017) destacam a existência de uma lacuna de comunicação entre governos e cidadãos. Uma lacuna que emerge da complexidade e extensão dos relatórios governamentais tradicionais. Como consequência, quando as autoridades não são efetivamente monitoradas por um “observador que tudo vê” como tal, a tendência é que não se comportem como o esperado, de forma que a qualidade do gerenciamento financeiro e de desempenho não é satisfatória (Eckersley *et al.*, 2014).

2.1 O Relatório Popular

Um novo modelo de relatório foi idealizado com o objetivo claro de endereçar essa lacuna. A iniciativa partiu do Conselho de Normas de Contabilidade Pública (GASB) e da Associação de Diretores Financeiros do Governo (GFOA) – ambos americanos (Lee, 2006). Com a publicação de *WhitePapers* (como o de Carpenter e Sharp em 1992) e a divulgação de um programa de premiação, os órgãos promoveram a implementação do que se chamou *Popular Annual Financial Report* - PAFR (Allison, 1995).

O Relatório Popular (RP) foi criado para complementar os relatórios tradicionais, e não os substituir, buscando aumentar o *accountability* (Lee, 2006). Trata-se de um material preparado pelos governos para os cidadãos, empresas e grupos sociais que querem entender informações gerais sobre as finanças e o desempenho das entidades públicas (Yusuf *et al.*, 2013). Políticos, funcionários públicos, mídia e outros grupos de interesse, como investidores e observatórios, também acreditam no RP como uma ferramenta útil para tomada de decisão (Cohen e Karatzimas, 2015).

Para isso, o RP busca simplificar as informações do Relatório Financeiro Anual Abrangente (*Comprehensive Annual Financial Report* - CAFR). Sendo “especificamente projetado para ser prontamente acessível e facilmente compreensível pelo público em geral e por outras partes interessadas sem experiência ou conhecimentos em finanças públicas” (GFOA, 2018).

Posteriormente, outros tipos de Relatórios Populares foram desenvolvidos. Todos os

modelos com algo em comum, o objetivo de comunicar informações financeiras aos cidadãos (Yusuf *et al.*, 2013). Exemplos: Guia de Cidadãos para o Orçamento, Orçamento Breve, Relatório Centrado no Cidadão, Resumo do Orçamento, Relatório de Gastos e Receitas, Relatório Anual, Relatório do Estado do Governo, Demonstrações Financeiras Resumidas e Relatório de Tendências Financeiras (tradução livre dos termos em inglês apresentados no artigo de Yusuf *et al.* de 2013). Nas grandes cidades e capitais dos EUA, os mais divulgados são o Relatório Centrado no Cidadão (*citizens' centric*, tradução livre) e o Relatório Financeiro Anual Popular (PAFR) (Yusuf *et al.*, 2013).

Os RPs, idealmente, têm características típicas. Ele deve ser elaborado com base nos relatórios tradicionais e os exigidos por lei, o conteúdo é resumido, de forma concisa, sem detalhes que aumentem a complexidade dos temas ou demandem uma análise mais aprofundada (Carpenter & Sharp, 1992; Cohen *et al.*, 2016). É importante que a referência aos relatórios oficiais fique clara para quem deseje buscar mais informações (Sharp, *et al.* 1998). Um RP deve conter textos e tabelas mais legíveis, que fazem uso de uma linguagem mais simples; com palavras e frases curtas, além de menos palavras técnicas ou de significado específico (Hermann, 2011). Terceiro, espera-se que o documento como um todo seja atrativo e didático, fazendo uso de ilustrações, figuras, gráficos e tabelas (Sharp, *et al.* 1998). Itens relevantes devem ser destacados (Hermann, 2011). Quarto, o relatório teria informações de interesse dos usuários (incluindo informações não financeiras, como quilômetros de via pavimentados, número de alunos atendidos pelas escolas, etc.) (Cohen e Karatzimas, 2015). Ou seja, deve ter foco naquilo que o cidadão médio deseja saber, informações que ele percebe como sendo úteis e interessantes (Oliveira, 2016). Quinto, as informações devem ser oportunas, sendo divulgadas frequentemente (no mínimo um relatório ao ano) (Carpenter & Sharp, 1992). Ainda, as informações devem ser críveis, na medida em que são elaboradas por uma fonte confiável, no sentido de que, informações que indiquem um desempenho ruim não sejam ocultadas ou disfarçadas (Hermann, 2011). Por fim, a prática da elaboração do RP encorajaria e facilitaria o feedback da população sobre o conteúdo e utilidade do relatório (Sharp, *et al.* 1998).

Considerando essas características, é possível avaliar se um RP é efetivo com base em quatro categorias (Yusuf e Jordan, 2012): informações financeiras relevantes; acessível e compreensível; visualmente interessante; e orientado para a comunidade. Especificamente sobre a compreensibilidade três aspectos são apontados: i) da presença de definições e explicações para termos técnicos e acrônimos; ii) do uso de jargão técnico; e iii) da escrita em

legibilidade de jornal - nível 7 ² (Yusuf e Jordan, 2012).

Com estas características normativas, o Relatório Popular em geral é visto como a “tradução dos relatórios tradicionais” (Paolo, Silvana & Valerio, 2016). Considerado uma ferramenta eficiente para divulgar informações financeiras de governos locais, àqueles que os artigos chamam de “cidadãos comuns”, pessoas sem experiência em contabilidade pública, – mesmo sem testes empíricos que validem esta expectativa de forma robusta.

Sobre os efeitos ou consequências dos RPs, quando os cidadãos podem acessar e entender os processos de orçamentação e de execução, é possível que eles participem de forma mais neutra na tomada de decisão governamental (Barbera et al., 2016). Assim, o Relatório Popular possibilita que esse processo aconteça em quatro estágios sequenciais: 1º) a capacidade de promover a transparência; 2º) a criação de um ambiente neutro; 3º) como consequência, haverá mais participação da sociedade, e por último 4º) impacto nos processos de tomada de decisão.

Em suma, o Relatório Popular seria uma ferramenta complementar, que ajuda os governos a serem mais transparentes e assim melhora as condições para o *accountability*. Da revisão de literatura, entende-se que os artigos colocam uma relação de causalidade direta clara: Relatório Popular é ‘igual’ a mais transparência e melhor *accountability*. Entende-se que essa consideração está baseada na suposição de que o Relatório, em sua versão Popular, melhora a compreensão do que se pretende comunicar. Talvez porque é bem consolidada a pressuposição de que um texto com frases e palavras mais curtas, usando um estilo de escrita mais simples e direto, seja mais legível e, portanto, mais fácil de ser compreendido. Ao parecer, uma conclusão amplamente aceita entre educadores, linguistas e literários, que estudam leitura e compreensão de textos por crianças e adultos, foi transposta ao contexto da contabilidade e finanças públicas.

Contudo, estudos na área de educação, especificamente sobre leitura, explicam que a compreensão que se depreende de um determinado texto depende de vários fatores (Dubay, 2004, p.1; Ganske e Fisher, 2009, p.285). Assim, quando se fala em melhorar a comunicação

² Em uma escala de 7 pontos, entre muito fácil, fácil, razoavelmente fácil, padrão, razoavelmente difícil, difícil e muito difícil, o nível 7 significa que o texto pode ser lido por crianças de 13 anos (ou sétimo ano completo) e adultos que têm um nível de proficiência de leitura próximo aos das crianças que completaram o sexto ano (e que deveriam ter 12 anos aproximadamente). Na escala de sete pontos, esse texto seria considerado ‘razoavelmente fácil’.

do desempenho fiscal de governos a pessoas sem conhecimento de contabilidade ou de finanças públicas, a questão pode não ser tão simples. O presente estudo busca contribuir com a literatura de Relatórios Populares aprofundando-se na interação entre Relatórios Populares e indivíduos (leitores). Para entender esse fenômeno, será usada a literatura de linguística (que trabalha o conceito de legibilidade) e de cognição (que explica como o texto é visualmente processado pelos indivíduos durante a leitura). Dessa forma, essa importante lacuna será trabalhada com o devido embasamento teórico e estrutura metodológica. Assim, a pesquisa analisa o nível de compreensão da informação de desempenho fiscal de governos focando não só no conteúdo, mas na interação entre leitor e leitura.

3. QUADRO TEÓRIO E DESENVOLVIMENTO DAS HIPÓTESES

3.1 Legibilidade e Compreensão

No começo do século 20, estudos na área de educação cresceram com foco na questão do aprendizado infantil (Wray e Janan, 2013). Uma parte da literatura se dedicou ao problema de seleção do material de leitura, partindo da ideia de que os textos deveriam ser adequados para os leitores (Dubay, 2004). Os resultados destes estudos influenciaram fortemente a estrutura e o conteúdo das grades curriculares nas escolas e universidades (Dubay, 2004). A ideia central era de que deveria haver um “casamento” entre leitores e textos, de forma que a legibilidade do texto fosse apropriada para a idade ou nível acadêmico (Janan, Wray & Pope, 2010).

Legibilidade pode ser definida como o que faz um determinado texto ser mais fácil de ler do que outros devido ao estilo da escrita (Klare, 1963). É quase intuitiva a relação de causalidade direta e positiva entre legibilidade e compreensão, como se assume nos estudos de transparência, *accountability* e relatórios populares. Contudo, a literatura em legibilidade aponta relações ainda mais complexas, incluindo a dependência de fatores intrínsecos ao leitor. Pode também ser definida como:

[...] a soma total (incluindo todas as interações) de todos aqueles elementos de um determinado material por escrito que afetam o sucesso que um grupo de leitores tem com o mesmo. O sucesso em questão é a velocidade de leitura, o quanto é entendido, e o quanto é percebido como interessante (Dale & Chall's, 1949 apud DuBay 2004, p. 3 – tradução livre).

Ou seja, as características físicas do texto importam, assim como a compreensibilidade que se obtém do mesmo; que depende também do leitor. Sobre os elementos textuais, o nível das palavras e das sentenças são as principais causas da dificuldade de leitura (Dubay, 2004, p.1). Quanto menos palavras desconhecidas (diferentes para o leitor) mais fácil será a leitura do texto; palavras mais longas tendem a ser mais difíceis; a estrutura das frases também afeta a dificuldade de leitura do texto, sendo frases mais longas mais difíceis (Chall, 1974, p. 157). Em suma, com base na literatura, destacam-se três características do texto: i) o comprimento médio das sentenças, ii) o comprimento das palavras e iii) o nível de dificuldade do significado das palavras (se são mais ou menos técnicos/ específicos).

Esses fatores já foram exaustivamente estudados e discutidos, e são, até hoje, amplamente

aceitos. Todavia, ainda não há um consenso claro sobre qual seria a melhor forma de mensurar esses fatores para a realização de experimentos e para aplicação das teorias nos problemas práticos. Uma forma muito popular de se operacionalizar essas questões é o uso de fórmulas de legibilidade. A partir do número de letras, sílabas, palavras, comprimento das sentenças e de outras características do texto, as formulas buscam medir o nível de dificuldade de um determinado texto. A fórmula de Dale-Chall (1948) por exemplo tem como resultado um índice que aponta o nível acadêmico mínimo para o qual o texto seria apropriado.³⁴

$$\text{Dale-Chall Score (1948)} = 0.1579 \left(\frac{\text{palavras difíceis}}{\text{palavras}} \times 100 \right) + 0.0496 \left(\frac{\text{palavras}}{\text{sentenças}} \right)$$

Fonte: Dale-Chall (1948)

O que Dale e Chall propuseram há 70 anos atrás segue sendo aplicado hoje: as fórmulas que medem legibilidade são instrumentos importantes na criação e adaptação de textos para audiências específicas. Com o tempo, pesquisadores das áreas de educação desenvolveram várias fórmulas, considerando diversos fatores. Com algumas limitações, elas podem medir de forma rápida e objetiva a dificuldade de leitura de um texto (Dubay, 2004, p.1 e 2). Ademais, com o avanço da tecnologia, criaram-se softwares com essa finalidade, o que colaborou para que as fórmulas se tornassem ainda mais populares (Dubay, 2004).

Para ilustrar, a tabela 1 lista seis das fórmulas muito utilizadas na língua inglesa. Cada uma é calculada levando em consideração fatores que determinam a legibilidade dos textos. Entre eles o comprimento das palavras, o comprimento das sentenças, o nível de complexidade das palavras, as palavras não-familiares (específicas) ou difíceis e as palavras polissilábicas.

³ Para mensurar a dificuldade do vocabulário, Chall (1974) propõe uma lista de palavras fáceis (em inglês).

⁴ O resultado mais baixo do índice equivale ao 5º ano do ensino fundamental brasileiro (crianças de 10 anos) e vai até o ensino universitário.

Tabela 1 – As formulas tradicionais de legibilidade

Fórmula	Variáveis preditivas				
	Comprimento das palavras	Palavras por anos de estudos	Comprimento das sentenças	Palavras não familiares /difíceis	Palavras polissílabas
FOG			*		*
Spache			*	*	
SMOG					*
Flesch-kincaid	*		*		
Dale-Chall			*	*	
ATOS	*	*	*		
ARI	*		*		

Fonte: adaptado de Janan & Wray (2012, p. 7.)

Outra fórmula que ficou muito conhecida (não consta na tabela) é o *Automated Readability Index* (ARI) (Senter e Smith, 1967). O ARI é calculado com base em uma metodologia diferente, que não depende de softwares avançados de contagem de sílabas (mais populares na língua inglesa). O ARI usa a proporção de palavras e sentenças difíceis, palavras com mais letras são consideradas mais difíceis (número de letras por palavra) e frases com mais palavras também (número de palavras por sentença) (Senter e Smith, 1967). Sua popularidade se deve ao fato de poder ser utilizado em diversos idiomas sem dificuldade, sem a necessidade de um software específico que saiba separar as sílabas (Dubay, 2004). E os resultados são muito semelhantes aos apontados por outras fórmulas mais já consolidadas pela literatura (Senter e Smith, 1967).

$$\text{Automated Readability Index (ARI)} = 4.71 \left(\frac{\text{caracteres}}{\text{palavras}} \right) + 0.5 \left(\frac{\text{palavras}}{\text{sentenças}} \right) - 21.43$$

Fonte: Senter e Smith (1967).

Apesar de serem muito utilizadas, as fórmulas em geral sempre foram criticadas; sua limitação consiste no fato de que não é possível captar a dificuldade que decorre do subjetivismo do texto (Dubay, 2004, p.2). As fórmulas se limitam às características físicas do texto, não contemplando a complexidade da narrativa (Ganske e Fisher, 2009, p.285). Essa complexidade emerge da interação entre leitor e leitura, e a legibilidade depende desse resultado (DuBay, 2004, p.7). Essa interação claramente também é influenciada por características do leitor. Algumas delas são: i) as capacidades físicas da pessoa; ii) as habilidades de leitura; iii) o engajamento ou motivação prévia; iv) conhecimento prévio e, v) gênero (Wray e Janan, 2013). Em suma, da literatura de legibilidade, entendemos que existem vários fatores que tornam um texto mais fácil ou difícil, sendo que eles não dependem apenas do texto, mas também do leitor e da sua interação.

3.2 O Movimento Ocular durante a Leitura

“Compreensão pode ser vista como o produto do desenvolvimento e da coordenação de várias habilidades necessárias para a leitura, incluindo o reconhecimento de palavras, a fluência de leitura, o processamento da sintática (estrutura das frases) e o conhecimento do significado das palavras do texto.” (Rayner et al., 2006 – tradução livre).

Quando se identifica um índice baixo de compreensão com base em testes ou perguntas, não se sabe exatamente que tipo de dificuldade ou de problema causou esse resultado (Rayner *et al.*, 2006). Por exemplo, não se sabe se foi a falta de conhecimento do significado de algumas palavras, ou se foram frases que estavam mal estruturadas. Por isso o monitoramento do movimento ocular durante a leitura contribui tanto para os estudos de sobre legibilidade e compreensão (Rayner *et al.*, 2006). O uso dos registros do olhar possibilita uma análise crítica sobre os processos cognitivos que ocorrem durante a leitura (Rayner, 1998). O mesmo ocorre sobre a forma como extraímos informações de um texto, e como integramos essas informações para interpretá-las (Clifton *et al.*, 2007 p.3).

Como a capacidade cognitiva do ser humano de processar o que se vê é limitada, é a atenção visual que seleciona o que será processado e direciona nosso olhar (Duchowski, 2007, p.4). Assim, os olhos tem que se mover para colocar em foco um ponto específico do nosso campo de visão, para ver com uma maior resolução o que quer que esteja no centro do nosso olhar (Duchowski, 2007, p.4). Por isso rastrear o caminho percorrido pelos olhos pode elucidar a que exatamente o observador dedicou atenção (Duchowski, 2007, p.3). Uma das formas de se entender esse caminho é pelo uso de métricas geradas pelos softwares dos fabricantes de equipamentos de *eye-tracking*.

Com base no fato de que a variabilidade das métricas reflete o processamento cognitivo online, o registro do movimento ocular sustenta, entre outros, um corpo expressivo de literatura que se dedica a entender os processos cognitivos que ocorrem especificamente durante a leitura (Rayner, 1998). O movimento ocular é considerado um reflexo que evidencia de que existe alguma dificuldade durante a leitura, ou ainda de que se obteve um baixo nível de compreensão (Rayner et al., 2006). Foi identificado, por exemplo, que certas propriedades das palavras influenciam o tempo de fixação (Rayner, 1998). É amplamente aceito que o tempo de processamento de um determinado trecho de texto pode ser mensurado pelo tempo de leitura acumulado da região, calculado pela soma das fixações sobre a região (Rayner, 1998).

Vários autores já sugeriram diversas formas de calcular o tempo de processamento de uma palavra, porém não há um consenso sobre qual seria o melhor e mais aceito. Por exemplo, sabe-se que podem haver efeitos *spillover*, quando o tempo de processamento de uma palavra é refletido nas fixações feitas nas palavras subsequentes (Rayner, 1998; Rayner, Sereno 1989). Isso porque, durante a leitura, algumas palavras são processadas mesmo quando não há uma fixação sobre a palavra (Rayner, 1998; Fisher and Shebilske, 1985). O movimento ocular pode ser influenciado também pela frequência da palavra alvo no texto; outro fator é a previsibilidade da presença da palavra alvo (Rayner, 1998: outros).

Enfim, o processamento cognitivo das palavras, ou morfológico como é chamado, é um evento complexo, que pode inclusive ocorrer com sobreposições (Rayner, 1998). Por isso o uso de qualquer medida de tempo de processamento por palavra é uma mera aproximação do processamento cognitivo (Rayner, 1998). Uma solução encontrada pelos pesquisadores é a adoção de alvos de análise no texto, sobre o qual é feita uma análise cuidadosa, usando diversas medidas (como duração da primeira fixação, tempo total de fixação, probabilidade de fixação em determinada palavra, número de fixações, duração das sacadas e efeitos de *spillover*) (Rayner, 1998).

Mais recentemente, os estudos têm se dedicado a entender como o nível de dificuldade geral de um texto, do discurso que ele carrega, afeta o movimento ocular. De acordo com os modelos *E-Z reader model* (Pollatsek *et al.*, 2006) e o *SWIFT model* (Engbert *et al.*, 2005), o movimento ocular é dirigido pelo processamento léxico. De acordo com esses modelos, o que direciona o movimento ocular é a necessidade de reconhecimento das palavras. Enfim, existe uma vasta literatura sobre os movimentos oculares durante a leitura, sendo que vários padrões de comportamento do olhar já foram identificados e testados em diferentes laboratórios pelo mundo (Rayner, 1998). Citaremos os que serão utilizados na presente pesquisa, que estão relacionados com a observação das dificuldades de leitura e da compreensão⁵. Normalmente se assume que, quando um texto é mais difícil, os leitores fazem fixações mais longas, sacadas mais curtas e mais regressões (Rayner, 2006). Esses padrões podem ser observados pelas

⁵ Ressalta-se que os pressupostos assumidos estão relacionados à leitura em silêncio, que é diferente da leitura em voz alta (Rayner, 1998).

métricas geradas pelo *eye-tracker*, entre elas o número de fixações, a duração média das fixações, o número de visitas, etc.

3.3 Hipóteses de Pesquisa

Com base nesse quadro teórico que objetiva explicar o comportamento ocular, a presente pesquisa propõe estudar o quanto as soluções propostas pelos Relatórios Populares são eficientes na entrega de maiores níveis de compreensão. Mais especificamente, se a maior legibilidade (sintática e léxica) e a ausência de jargões realmente leva a um maior nível de compreensibilidade dos relatórios. Para responder a esse questionamento as seguintes hipóteses de pesquisa serão testadas:

H_1 : Quanto menor a sofisticação léxica (comprimento das palavras) e complexidade sintática (comprimento das frases), maior será a compreensão.

H_2 : Quanto menor o uso de palavras difíceis (significado complexo, específico ou técnico), maior será a compreensão.

H_3 : Quando termos técnicos são substituídos por outros mais comuns, haverá indícios de maior interesse e esforço cognitivo durante a leitura.

4. METODOLOGIA

Para observar se a legibilidade dos relatórios públicos afeta a compreensão e a atenção visual dos indivíduos, foi desenvolvido um experimento composto de três testes ou tarefas. Cada um deles com uma manipulação diferente para testar as hipóteses H1, H2 e H3 já apresentadas. O que diferencia e enriquece a metodologia dessa pesquisa é a utilização de um equipamento de rastreamento ocular. O *Eye-tracker*, como é conhecido, forneceu evidências sobre os processos cognitivos que ocorriam enquanto os participantes observavam o conteúdo apresentado. O procedimento do experimento foi explicado detalhadamente no Apêndice A.

O protocolo do experimento foi aprovado no Comitê de Ética em Pesquisa (CEP) sediado na Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, da Universidade de São Paulo. O ofício de aprovação foi emitido no dia 28 de novembro de 2018 - vide Anexo 1.

A inclusão do equipamento de rastreamento ocular, ou *Eye-tracker* como é conhecido, é importante para a pesquisa pois é uma metodologia que evita vieses nas respostas por reflexividade dos participantes. Quando é o participante quem declara aspectos ou fatores de si mesmo, as respostas podem ser distorcidas pois as pessoas tendem a mostrar o que se considera socialmente desejável (Meibner e Oll, 2017). Por isso a presente pesquisa adota uma abordagem diferente, pouco utilizada no campo de contabilidade pública (Grimmelikhuijsen *et al.*, 2017).

Contudo, é preciso reconhecer que mesmo os métodos experimentais não são perfeitos, e poderiam haver outros tipos de viés, como o ‘efeito demanda’ (Charness, Gneezy e Kuhn, 2012). Nesse caso, o participante tende a se comportar da forma como se percebe que é desejável pela pesquisa. Como o presente experimento envolve atividades que dependem mais de habilidades do que de preferências, um viés sistemático desse tipo é improvável (Charness *et al.*, 2012) - os testes não se referem aos gostos ou às preferências dos indivíduos, mas sim buscam observar sua capacidade de compreensão e seu comportamento ocular ao ler relatórios de governos. Assim, considera-se que a metodologia escolhida é um ponto forte da pesquisa.

A escolha da metodologia também se deve a necessidade de amenizar possíveis vieses por efeito de leniência, reflexividade ou halo. Com o *eye-tracker*, os dados são coletados de forma ininterrupta e em tempo real, enquanto o participante realiza a atividade; por isso não são contaminados pelos efeitos de demanda por memória, processos de articulação ou estratégias mentais conscientes (Rayner, 2006). Outra vantagem é a possibilidade de inferir sobre os momentos exatos em que há maior ou menor dificuldade de compreensão do texto, com base em métricas, como, por exemplo, a probabilidade ou a duração das fixações (Rayner, 2006).

4.1 O desenho do experimento

Os métodos experimentais são caracterizados pela capacidade de medir sistematicamente o resultado de duas (ou mais) condições, sendo uma a condição de controle e a outra a condição de tratamento (Stausberg & Engler, 2013, p.162). Quando o controle experimental é absoluto, é possível comparar os resultados e afirmar que todas as diferenças ocorrem devido ao fator de condição; é o que se chama de ‘experimento perfeito’ (Stausberg & Engler, 2013, p.162).

Todo procedimento experimental tem pelo menos uma variável dependente - *outcome* final do experimento e representativa do fenômeno a ser explicado -; e uma (ou mais) variáveis independentes – aquela que é manipulável, introduzida por uma intervenção (Stausberg e Engler, 2013, p.164). É comum que haja variáveis de controle, também chamadas de covariantes. São fatores importantes que impactam a variável dependente, mas não compõem o foco principal da pesquisa e/ou não são manipuláveis (Stausberg e Engler, 2013, p.164).

Tanto experimentos como quase-experimentos podem ocorrer entre sujeitos (*between subjects*), ou intrassujeitos (*within-subjects*) (Charness et al., 2012). Entre sujeitos, a correspondência entre todos os fatores concebíveis é impossível; ou seja, é impossível encontrar dois grupos de pessoas que tenham exatamente as mesmas características, incluindo gostos, preferências, crenças, etc. Nesse formato, a probabilidade de distribuição aleatória desses fatores deve ser baixa o suficiente, de forma que se possa inferir que as diferenças nos resultados entre os grupos se dão devido à manipulação experimental (Lourenço, Hayashi & Almeida, 2009). Isso significa que os participantes devem ser designados aos grupos de forma totalmente aleatória (Charness et al., 2012). Ressalta-se ainda que, idealmente, para fazer as análises entre sujeitos por hipótese, as manipulações devem ser feitas na mesma ordem para todos (Charness et al., 2012). No desenho intrassujeitos o procedimento de seleção aleatória não se faz necessário. Isso porque as comparações são feitas acompanhando o desempenho do mesmo participante sob diferentes condições: o sujeito é seu próprio controle (Lourenço Hayashi & Almeida, 2009).

A pesquisa adotou o delineamento mais básico, de duas condições de análise: linha de base e intervenção. Em todos os três testes, na condição de base são apresentados os textos de baixa legibilidade. Na intervenção apresentam-se textos com alta legibilidade. O formato adotado foi entre sujeitos, e os testes sempre ocorrem sempre na mesma ordem. Buscando o

melhor controle experimental possível, a amostra selecionada pertence a um mesmo grupo de alunos, a fim de que haja maior correspondência entre os fatores que diferenciam os indivíduos, principalmente no que se refere a capacidade cognitiva. Mais detalhes sobre os critérios e seleção da amostra são descritos na seção 5.2 Amostragem.

De forma resumida, no teste 1, manipula-se o nível de legibilidade alterando o número de palavras por frase e comprimento das palavras. Nos testes 2 e 3 a legibilidade manipulada foi devido à presença de termos técnicos da área de contabilidade e finanças públicas. As variáveis dependentes são as métricas geradas pelo software do *eye-tracker* e os resultados obtidos nos testes de compreensão. As variáveis de controles são gênero, idade, *background* acadêmico e os resultados do Testes de Reflexão Cognitiva.

Após a coleta, os dados foram processados usando o software do fabricante do aparelho. Nessa etapa o pesquisador configura para cada lâmina do experimento, ou estrato visual de relatório apresentado, as áreas de atenção a serem monitoradas, as chamadas ‘AOIs’. A partir de cada AOI, o aparelho monitora a atenção visual do respondente e gera as métricas, além de relatórios do tipo ‘mapas de calor’ (veja Apêndice C). Depois, os dados foram importados e padronizados constituindo uma base com 50 linhas (n da coleta = 50) e 1329 observações. As análises, os testes estatísticos e os gráficos foram feitos usando o software StataSE 12.

4.2 O Equipamento de rastreamento ocular (*Eye-Tracker*)

Conforme citado, foi utilizado um dispositivo que acompanha e registra os movimentos oculares: o *Tobii Eye Tracker* modelo X1L. O rastreamento ocular de alta precisão ocorre pelo efeito de dois iluminadores óticos: um foca no meio e o outro fora do eixo ótico. Dessa forma cria-se um efeito de claro-escuro na pupila que possibilita determinar a posição exata dos olhos (TOBII, 2013).

De acordo com TOBII (2013), as especificações do modelo são as seguintes:

- Sistema de câmeras duplas, com rastreamento de ambos os olhos simultaneamente;
- Seleção automática de claro e escuro para registro dos movimentos da pupila;
- Captura média de 30 amostras por segundo de cada olho (30 HZ) e;
- Tempo de recuperação do rastreamento após uma piscada de olhos de 10 ms.

4.3 As métricas geradas pelo *Eye-tracker*

Ao registrar o movimento ocular, a cada minuto de gravação o *Eye-tracker* gera uma quantidade enorme de dados. É um software do próprio fabricante do equipamento que transforma toda essa informação em métricas, o *Tobii Studio Professional 3.2*. Nesse estudo serão usadas 7 métricas para entender o comportamento ocular, que estão entre as mais usadas. A tabela 2 a seguir resume o nome original da métrica em inglês, o nome para o qual a traduzimos e a explicação do que ela cada uma delas mede, de acordo com o manual do usuário do software citado (Tobii AB, 2016). Na última coluna apresenta-se as letras usadas nos nomes das variáveis.

Tabela 2 – Métricas do *Eye-tracker*

Métrica em Inglês	Tradução livre	Descrição da métrica	Unidade	Variável
Fixation Duration	Tempo Médio de Fixação	Duração média de todas as fixações registradas dentro de uma AI, ou dentro de todas as AIs de um grupo de AIs.	segundos	fm
Total Fixation Duration	Tempo Total de Fixação	Soma de todas as fixações registradas dentro de uma AI, ou dentro de todas as AIs de um grupo de AIs.	segundos	ft
Fixation Count	Número de Fixações	Número de fixações registradas dentro de uma AI, ou dentro de todas as AIs de um grupo de AIs.	contagem	fc
Visit Duration	Duração Média da Visita	Duração média de todas as visitas registradas dentro de uma AI, ou dentro de todas as AIs de um grupo de AIs.	segundos	vm
Total Visit Duration	Duração Total da Visita	Soma de todas as visitas registradas dentro de uma AI, ou dentro de todas as AIs de um grupo de AIs.	segundos	vt
Visit Count	Número de Visitas	Número de visitas registradas dentro de uma AI, ou dentro de todas as AIs de um grupo de AIs.	contagem	vc
Time to Mouse Click	Tempo até o clique	Duração do teste ou tempo até o participar dar o clique no mouse.	segundos	et

Fonte: elaborado pelo autor.

4.4 Jovens Adultos e o “Cidadão comum”

Em ambas literaturas, sobre Relatórios Populares e transparência, “o cidadão comum” ou “os cidadãos” de forma mais genérica são mencionados. Os cidadãos podem ser entendidos como aqueles que pagam as contas do governo, que deveriam, de acordo com o modelo de vigilância aparentemente vigente ao redor do mundo, monitorar o desempenho dos governantes e reagir por meio do voto. Inclusive nos pronunciamentos contábeis internacionais do setor público (IPSAS), os cidadãos são vistos como usuários da informação contábil. Seu papel

parece claro, porém raras vezes o que o termo “cidadãos” representa é explicado em detalhe.

Em alguns casos, principalmente nos estudos sobre Relatórios Populares, é explicitado apenas que os cidadãos em questão são indivíduos comuns, não-especialistas em contabilidade ou finanças públicas (e.g. Carpenter & Sharp, 1992; Yusuf e Jordan, 2012; Barbera et al., 2016). Entende-se que essa breve explicação não necessariamente se refere ao status social ou à idade dos indivíduos, mas pode estar associado à atividade profissional que ele desempenha ou já desempenhou, pois é o que levaria a uma maior ou menor familiaridade com o conhecimento técnico necessário para a compreensão de relatórios financeiros. Por exemplo, técnicos em contabilidade podem ter mais facilidade em relação à doutores em medicina, artes, história ou física. Outra suposição é na maioria das vezes não exercem cargos na administração pública ou posições em conselhos municipais que permitam o desenvolvimento das habilidades e o conhecimento necessário para realizar as tais análises de desempenho dos governos.

Está fora do escopo dessa pesquisa o entendimento do que é este “cidadão comum” em si, mas considera-se um ponto importantíssimo para o tema aqui trabalhado, a ser discutido pelas pesquisas futuras no tema. Nesta pesquisa, o cidadão comum é entendido como aquele que não possui experiência prévia ou conhecimento formal estabelecido (i) na dinâmica de reconhecimento, mensuração e registro contábil (ou elaboração de relatórios financeiros) e (ii) finanças públicas. Como analogia, em um espectro, é como se esse cidadão comum estivesse em um extremo, enquanto especialistas em contabilidade e finanças públicas se encontram no outro extremo. Essa divisão é uma abstração, dado que existem diversos níveis de especialização neste contínuo.

Além disto, a revisão da literatura sobre compreensão durante a leitura leva a outro ponto importante, de que mesmo entre o grupo dos ditos cidadãos comuns, no extremo de menor especialização no tema, coexistem diferentes níveis de habilidades cognitivas. Por exemplo, capacidade da memória de trabalho ou a habilidade de usar o sistema cognitivo analítico-racional no lugar do sistema intuitivo. Foi feita a opção por selecionar um estrato da população de “cidadãos comuns” (sem conhecimentos específicos ou experiência de trabalho em contabilidade ou finanças públicas) com um alto nível de habilidades cognitivas em relação à média nacional.

A seleção de alunos do 1º ano dos cursos de administração e economia em uma Universidade pública oferece jovens adultos com níveis altos e similares de habilidades cognitivas. Esse pressuposto é feito com base no fato de que todos foram bem-sucedidos em um vestibular altamente competitivo alguns meses antes da coleta. Em sua maioria (quase 86%

da amostra) tem entre 18 e 21 anos de idade. A Figura 1 abaixo apresenta um gráfico com a distribuição dos participantes de acordo com sua idade.

Figura 1 – Distribuição dos participantes de acordo com a idade (anos).

Fonte: elaborado pelo autor.

Esses jovens, nascidos entre 1997 e 2000, pertencem a chamada geração Z, também conhecida como “Centennials” (Geck, 2007, p. 235). A Geração Z se refere aos indivíduos nascidos após 1990, sendo uma forte característica desse grupo o fato de terem convivido desde muito cedo com a Internet como ela é hoje. Apesar de terem muita facilidade com as ferramentas mais populares de busca de informações na internet, é comum que lhes falte habilidades para avaliar os conteúdos encontrados (Geck, 2007, p. 235). Esses indivíduos conviveram com um mundo tecnológico completamente diferente da geração anterior, e dificilmente buscam outras fontes de informação que não sejam a própria internet (Geração Y ou “millennials”) (Geck, 2007, p. 235).

4.5 Amostragem

O processo de seleção da amostra garante o controle experimental principalmente por minimizar possíveis efeitos por discrepância sistemática entre os grupos. A capacidade cognitiva dos participantes e o *background*, fatores que influenciam a capacidade de leitura e compreensão (Wray e Janan, 2013), foram controladas no processo de seleção da amostra.

Como explicado na seção anterior, assume-se que os participantes representam um estrato da população que tem uma alta capacidade cognitiva. Contudo, sabe-se que os participantes não possuem necessariamente uma capacidade cognitiva semelhante.

Já a influência do *background* foi controlada selecionando jovens com nível similar de exposição prévia à: (i) dinâmica de reconhecimento, mensuração e registro contábil, ou elaboração de relatórios financeiros; e (ii) finanças públicas. A amostra de jovens adultos inclui alunos de primeiro ano dos cursos de Ciências Econômicas e Administração de Empresas de uma das unidades de ensino de uma universidade pública brasileira. À época da coleta, os alunos de administração já haviam concluído a disciplina Contabilidade Empresarial e os de Economia, a de Contabilidade Social. Portanto, a amostra é de jovens adultos com conhecimentos básicos de contabilidade geral - não de contabilidade pública.

Para reforçar esse controle experimental foi feito ainda o teste de calibração dos jargões, explicado em detalhes na seção 4.6. Assim, é pouco provável que haja uma influência sistêmica nos resultados do experimento devido a um maior conhecimento prévio do significado de jargões ou de lógicas de interpretação da contabilidade pública por parte da amostra. Ainda, ao selecionar alunos de 1º ano, trabalha-se com alunos com nível similar de preparação profissional, antes de passarem pelo desenvolvimento de todo o curso de graduação.

As listas de presença dos alunos serviram de base para organizar a lista de potenciais respondentes. A população amostral totalizou 141 alunos, com idade entre 17 e 27 anos, como mostra a Tabela 3. A amostra convidada a participar da coleta foi aleatoriamente definida, com controle de gênero. Uma nova lista foi gerada contendo 20 mulheres e 20 homens de cada turma, selecionados por sorteio. Esses alunos foram então abordados por e-mail e também por telefone. A maioria deles aceitou o convite para participar da pesquisa.

Tabela 3 – População referência e amostra

	Mulheres	Homens	População amostral	Mulheres	Homens	Amostra
Administração Diurno	18	33	51	10	12	22
Administração Noturno	14	29	43	6	8	14
Economia Noturno	13	34	47	5	9	14
Total	45	96	141	21	29	50
Média de idade				19.28	19.44	19.38

Fonte: elaborado pelo autor.

4.6 Identificação e validação dos jargões na área (termos técnicos).

Para a seleção e validação dos jargões da contabilidade pública a serem utilizados nos testes, foi feito um processo de validação dos termos com os indivíduos da própria população amostral. A ideia por trás dessa medida era poder realizar um tipo de calibração do nível de dificuldade dos jargões, afim de que se tivesse uma maior validade experimental interna.

Primeiro, usando um relatório de 2017 da Secretaria da Fazenda do Estado da Bahia – a Prestação de Contas Anual Simplificada – 152 termos técnicos (contendo uma ou mais palavras) foram identificados. Os termos representam jargões pois carregam significados muito específicos ou técnicos do campo da contabilidade e das finanças públicas, sendo encontrados no texto e nas tabelas do relatório. Depois, os termos foram listados em uma planilha e pré-classificados pelos autores e especialistas na área. Foram criadas três categorias, fácil, médio e difícil, com base no nível de dificuldade do seu significado. Os critérios utilizados para a classificação foram sugeridos pelos autores e corroborados pelos especialistas, como mostra a Tabela 4 abaixo.

Dessa lista, 30 termos foram escolhidos para serem validados pela população amostral. Ou seja, essa validação foi feita com o mesmo grupo de respondentes que dias depois foi convidado a participar do experimento (mais detalhes na seção 4.4 Amostragem). Foram escolhidos 10 termos de cada nível (fácil, médio e difícil), selecionados de forma que fossem variados e pudessem posteriormente construir diferentes narrativas no experimento. Foram criadas então duas listas com 15 termos cada uma por meio de randomização.

As duas listas foram dispostas em forma de questionário usando uma plataforma online, para que os próprios alunos (que ainda não sabiam sobre a pesquisa) classificassem, de acordo com a sua percepção, o quanto eles tinham conhecimento do significado de cada um dos termos. Os alunos responderam o questionário em sala de aula, com seus próprios celulares, por meio

de um hiperlink compartilhado por e-mail.

Tabela 4 – Validação dos Jargões: Critérios para definição dos termos difíceis

Nível de dificuldade	Critérios para definição do nível de dificuldade	Jargões identificados
Fácil	compreensão quase que imediata dedutível pelo nome: não requer conhecimentos aprofundados ou definição prevista em lei	44
Médio	poderia ser deduzida pelo nome sem conhecimento porém implica algum raciocínio lógico necessita conhecimento básico de definições previstas em lei	76
Difícil	necessita conhecimento das leis e conceitos (definições extensas) termos mais técnicos da área de contabilidade, que se referem a operações específicas demanda raciocínio lógico mais avançado	32
Total		152

Fonte: elaborado pelo autor.

Assim, responderam a seguinte pergunta, em formato de escala Linkert de 4 pontos:

1. Para cada um dos termos a seguir, qual seu nível de familiaridade?

‘termo’ – por exemplo: “Renúncia de receita”

(1) Nenhuma (não conheço este termo)

(2) Pouco familiar (tenho algum conhecimento sobre o significado)

(3) Razoavelmente familiar (tenho conhecimento razoável sobre o significado)

(4) É familiar (sei o significado)

O objetivo da validação foi reduzir as chances de incluir termos considerados fáceis ou difíceis, cujo grau de dificuldade não estivesse alinhado com a dificuldade do próprio grupo de respondente. Assim, espera-se que a manipulação dos textos represente, como diz a literatura, significados com as quais o leitor tem maior ou menos familiaridade. O resultado da classificação dos jargões pode ser visto em duas tabelas que foram apresentadas no apêndice F.

4.7 O Desenho dos Testes 1, 2 e 3

Cada teste foi desenhado para testar uma das hipóteses apresentadas. Em todos eles as variáveis dependentes a serem analisadas incluem as sete métricas apresentadas na Tabela 2

(Tempo Médio de Fixação, Tempo Total de Fixação, Número de Fixações, Duração Média da Visita, Duração Total da Visita, Número de Visitas e Tempo até o clique). Os testes 1 e 2 incluem ainda 2 perguntas de compreensão com resposta em forma de múltipla escolha. Os três testes serão apresentados nas próximas seções com mais detalhes. Ressalta-se que as versões A sempre serão aquelas em que esperamos um maior nível de compreensibilidade e de facilidade de leitura, com modificações que remetem aos Relatórios Populares. As versões B do teste apresentam elementos típicos de relatórios tradicionais da contabilidade pública, onde espera-se que os participantes encontrem mais dificuldade durante a leitura e nas perguntas de compreensão. Simplificando: ‘A’ sempre se refere ao que se poderia chamar de “versão fácil”, e B à “versão difícil”. O material usado em todos os testes pode ser visualizado na íntegra no Apêndice B e C.

4.7.1 O Teste 1 – Legibilidade sintática e léxica dos textos (relatos de desempenho de governos)

O teste 1 foi desenhado para testar a hipótese de pesquisa H1. O objetivo é saber se a sofisticação léxica (comprimento das palavras) e a complexidade sintática (comprimento das frases) influencia o nível de compreensão dos leitores quando se trata de relatórios financeiros de governos. Aqui o participante se deparou com uma tela de fundo branco na qual havia um texto que falava sobre as entregas de uma secretaria de infraestrutura fictícia. O texto também falou sobre as finanças da prefeitura, relacionando um empréstimo e doações de empresários às obras realizadas naquele ano. A principal mensagem que os leitores poderiam extrair é a de que não haviam recursos públicos disponíveis para realização das obras, que foram realizadas com capital de terceiros.

Entre as versões A e B do teste, altera-se o comprimento de algumas palavras, que foram substituídas, e das frases, que foram reestruturadas. Na versão A, as frases e as palavras são mais curtas. O conteúdo informacional dos textos é o mesmo, mantendo-se o tom do discurso, para que nenhuma das versões levasse o leitor a interpretar a situação de forma diferente. Esse controle é importante, pois reforça a validade interna do experimento. Assim, da comparação dos resultados, é possível inferir que as diferenças se dão devido a manipulação, e não a outras características do texto.

A tabela abaixo mostra os índices de legibilidade calculados, segundo a fórmula do *Automated Readability Index* (ARI). Foram utilizadas as páginas web *Online-Utility.org*,

StoryToolz e *WebFX*⁶, que fazem o cálculo desse e de outras fórmulas de legibilidade online, gratuitamente. Cada um tem seu sistema de contagem e de cálculo automatizado, e de acordo com os próprios websites, a fórmula que se aplica é a mesma - de Senter e Smith (1967), citada previamente. Assim mesmo, há uma pequena diferença nos resultados. Por isso optou-se por utilizar a média simples entre os três valores, como mostra a Tabela 5. O resultado do índice é equivalente a idade mínima em anos que o leitor deveria ter para entender o texto (Senter e Smith, 1967). Assim, quanto menor o valor, maior a legibilidade.

Tabela 5 – *Automated Readability Index* (ARI) dos Testes 1 e 2

Teste	WEB FX	Online-Utility.org	StoryToolz	Média	Nº de jargões (classificação)	Nº de palavras
T1A (maior legib.)	8.6	9.96	7.1	8.55	3 (fácil)	155
T1B	18.5	19.91	18.6	19.00	3 (fácil)	176
T2A (maior legib.)	12.7	13.69	11.7	12.70	3 (fácil)	132
T2B	12.1	13.13	11.1	12.11	3 (difícil)	129

Fonte: elaborado pelo autor

No teste 1, para o texto A, o índice de legibilidade médio é 8.55. Assim, considerando as características físicas do texto, o estilo da escrita é adequado para crianças a partir dos 8 anos. Já o texto B seria recomendado para jovens adultos de 19 anos ou mais. Foram usados 3 jargões da contabilidade, todos classificados como de fácil compreensão na enquete de validação dos jargões, sendo os mesmos em ambas versões.

Após a leitura do teste, o participante responde 2 questões com respostas de múltipla escolha na própria tela do computador, usando o mouse. A primeira pergunta é mais simples e se refere a um fato que está evidente no texto, apenas uma resposta é correta. Essa pergunta é uma *proxy* para medir se o participante leu o texto com atenção. Basicamente, a tarefa cognitiva que o indivíduo teria que desempenhar para acertar essa questão era a de recordar o evento “obras nas escolas” e associá-lo a um tipo de recurso (questão completa e opções de resposta no APÊNDICE C). Entende-se que se trata de uma tarefa fácil, mesmo no texto com menor

⁶ Online-Utility.org - Utilities for Online Operating System: https://www.online-utility.org/english/readability_test_and_improve.jsp, WebFX Digital Marketing That Drives Results: <https://www.webfx.com/tools/read-able/check.php> e StoryToolz <https://storytoolz.com/readability>.

legibilidade, pois a informação foi apresentada, de forma explícita, nas últimas linhas do texto.

Já na segunda pergunta, o que se mede é a compreensão do texto como um todo. Trata-se de uma tarefa cognitiva mais desafiadora, que busca saber se o leitor entendeu um fato que reside nas entrelinhas do texto. Em outras palavras, trata-se de uma informação implícita, que se extrai da interpretação do texto. A resposta - de que as obras foram financiadas com recursos de terceiros - depende de outro trabalho cognitivo, de interpretação do texto.

Com base na literatura revisada, quem teve acesso ao texto facilitado (T1A) deveria ter um nível de compreensão melhor, e, portanto, o grupo que seguiu o “Caminho 1” também deveria acertar mais essa questão. Outra suposição que se pode fazer, a partir da literatura, é a de que os participantes que acertaram a questão 1 (t1q1 - proxy para atenção) deveriam acertar mais a questão 2 (t1q2 - proxy para compreensão).

Ainda com foco na Hipótese de pesquisa H_1 , além dos resultados das perguntas de atenção e compreensão, foram consideradas as métricas geradas pelo *eye-tracker*. Da literatura sobre movimento ocular durante a leitura, espera-se que o ritmo de leitura dos participantes seja maior em T1A devido à legibilidade do texto (em forma de frases e palavras mais curtas, medida pelo ARI apresentado na tabela 5). O texto mais legível deve ser mais fluido, facilitando não só a leitura como o entendimento. Esse fator se reflete nas métricas, que servirão como *proxies* da facilidade de leitura. Como já explicado na seção 3.2, espera-se que o grupo que teve acesso ao texto facilitado apresente menos fixações, fixações mais curtas, menos visitas e visitas mais curtas.

4.7.2 O Teste 2 – Jargões nos textos (relatos de desempenho de governos)

O Teste 2 foi desenhado com foco na Hipótese de pesquisa H_2 . Foram apresentados dois textos, ambos sobre o desempenho financeiro e operacional de um governo estadual, contando que a receita arrecadada naquele ano havia sido menor e que devido a isso a alguns dos serviços prestados pelo estado tiveram que ser suspensos. O texto cita também outra decisão do governo e o problema da previdência do estado. Se interpretado, é possível depreender que o governo estadual teve problemas devido à queda na arrecadação das receitas.

Diferente do teste 1, nas versões A e B, a estrutura do texto não muda. O nível de legibilidade inclusive foi usado como um controle, afim de garantir que, em questão de estrutura sintática e sofisticação léxica, os textos fossem o mais parecidos possível. A tabela 5

apresentada na última seção mostra que nas duas versões do texto a idade mínima recomendada para leitura seria de 12 anos. Além disso, o conteúdo informacional é o mesmo. Esse controle é importante, pois reforça a validade interna do experimento. Assim, é possível inferir com maior segurança sobre os efeitos da presença/ausência dos jargões, descartando possíveis interferências nos resultados do experimento devido à outras características do texto.

Assim, a legibilidade é manipulada pela presença de termos técnicos ou de significado específico. No caso, os termos são os jargões da área de contabilidade e finanças públicas. Ao longo da versão B do texto (T1B) aparecem, dispersos, 3 jargões previamente classificados como de difícil compreensão no processo de validação dos jargões (seção 4.6 Validação dos Jargões). Os termos usados são apresentados mais adiante, na seção 5.1.2 Análise do Teste 2. Já a versão A do texto (T2A), é como se fosse uma versão “traduzida”, sem os jargões da contabilidade. No lugar dos termos, foram usadas outras palavras, menos técnicas, mas que ainda transmitem o mesmo significado.

Com base na literatura revisada, supõe-se que a compreensão do texto seria maior entre os participantes que leem essa versão simplificada, sem os termos técnicos da contabilidade. Foram comparadas as métricas geradas pelo *eye-tracker*, que serão usadas como *proxy* da facilidade de leitura. Nesse teste em especial podemos comparar as métricas geradas para as AOIs que cobrem exatamente os termos contábeis e os termos alternativos (com palavras menos técnicas).

Como no teste 1, a primeira pergunta é uma *proxy* para medir a atenção do participante, enquanto a pergunta 2 mede sua interpretação. Quem teve acesso ao texto sem jargão (T2A) deveria ter um nível de compreensão melhor. Portanto, espera-se que a taxa de acerto de ambas perguntas seja maior entre os participantes que acessaram o teste T2A, ou seja, *t1q1* e *t1q2* serão maiores para o grupo que fez o Caminho 1. De forma análoga ao Teste 1, supõe-se que os participantes que acertaram a pergunta de atenção (*t1q1*) deveriam acertar mais a pergunta de interpretação (*t1q2*).

4.7.3 O Teste 3 – Jargões em tabelas (demonstrações contábeis de governos)

O Teste 3 foi feito com base na hipótese de pesquisa H_3 e busca averiguar se a apresentação de conteúdo contábil e financeiro, em forma de tabela, usando palavras menos técnicas, altera de alguma forma o padrão do comportamento visual dos leitores. Pra isso foi apresentado um

Balanco Patrimonial Geral de um Município fictício, contendo 5 contas do ativo e 4 do passivo, além das somas de circulante e não circulante em duas colunas, ano atual e ano anterior. Para direcionar o trabalho cognitivo dos participantes, a pergunta “Qual a situação financeira da prefeitura?” foi deixada logo abaixo da tabela. Entende-se que ao ler a pergunta, os participantes tentariam respondê-la em silêncio, olhando para os termos e valores na busca de uma resposta.

A tabela foi criada pelos autores com base nas demonstrações contábeis divulgadas pela prefeitura de Ribeirão Preto. A versão B do teste (T3B) usa alguns termos que são exatamente iguais aos dessas demonstrações. As linhas representam contas comuns, descritas por vocabulário técnico, tradicional da contabilidade, de acordo com as normas brasileiras de contabilidade pública. Para a versão A, que é a manipulada, os termos técnicos foram substituídos por palavras menos estranhas, menos técnicas, com maior chance de serem consideradas como familiares pelo leitor.

Um maior interesse em entender a situação deve ser refletido no tempo que o participante passa olhando para a tela, supõe-se que o tempo para o clique no mouse deve ser maior na versão A. Além disso, nessa mesma versão, de acordo com a literatura em legibilidade, o participante deveria gastar menos tempo olhando para as palavras da tabela, pois seriam mais facilmente entendidas.

4.8 Aplicação do experimento

Como explicado, o experimento consiste de uma única sessão de aproximadamente 20 minutos em que se apresentaram 3 testes no formato entre sujeitos. Os testes foram chamados de Teste 1, Teste 2 e Teste 3 e foram abreviados para T1, T2 e T3. Soma-se letra “A” para identificar a versão fácil, na qual há manipulação para haver alto nível de legibilidade. A letra “B” denomina as versões mais difíceis, com um baixo nível legibilidade. A todos os participantes apresentou-se primeiro T1, seguido de T2 e por último T3. Havia duas sequências de experimento possíveis, que chamamos de “Caminho 1” e “Caminho 2”. Os participantes que realizam o Caminho 1 fizeram a sequência T1A, T2B e T3A, enquanto os outros realizaram T1B, T2A e T3B. A tabela 6 apresenta um resumo das coletas realizadas, com gênero, idade média dos respondentes e caminho. O protocolo de coleta é explicado detalhadamente no Anexo 1.

Tabela 6 – Caminhos por gênero dos respondentes

	Homens	Mulheres	Total
Caminho 1 (A B A)	15	9	24
Caminho 2 (B A B)	14	12	26
Total	29	21	50

Fonte: Elaborado pelo autor

4.9 Estatística descritiva do experimento

Tabela 7 – Descritiva da compreensão dos experimentos e CRT

Variável	Sigla	Descrição	n	Média	D.p.	Mín.	Máx.
Teste 1							
Questão 1	t1q1	Nível de atenção (1/0)	50	0.72	0.45	0	1
Questão 2	t1q2	Nível de compreensão (1/0)	50	0.6	0.49	0	1
Teste 2							
Questão 1	t2q1	Nível de atenção (1/0)	50	0.9	0.3	0	1
Questão 2	t2q2	Nível de compreensão (1/0)	50	0.9	0.3	0	1
Teste de Reflexão Cognitiva (CRT)							
Questão 1	crt1	Capacidade cognitiva (1/0)	50	0.4	0.49	0	1
Questão 2	crt2	Capacidade cognitiva (1/0)	50	0.34	0.47	0	1
Questão 3	crt3	Capacidade cognitiva (1/0)	50	0.92	0.27	0	1
Soma	crt_t	Soma de 3 CRT	50	1.66	0.89	0	3
Questão 4	crt4	Percepção de acerto CRT (0-4)	50	2.34	0.68	1	3

Fonte: Elaborado pelo autor

Tabela 8 – Métricas do ET para os 3 testes

Métrica (sigla)		Teste 1		Teste 2		Teste 3	
		A n = 24	B n = 26	A n = 26	B n = 24	A n = 24	B n = 26
Número de	média	183.50	181.61	156.38	189.63	172.875	158.54
fixações (fc)	(d.p.)	(64.93)	(58.49)	(69.11)	(95.82)	(90.46)	(67.82)
Tempo médio de	média	0.38	0.37	0.37	0.36	0.377083	0.4
fixação (fm)	(d.p.)	(0.06)	(0.05)	(0.07)	(0.06)	(0.09)	(0.046)
Tempo total de	média	71.36	67.86	59.64	71.41	68.0225	63.40
fixação (ft)	(d.p.)	(23.62)	(23.66)	(28.36)	(35.44)	(37.87)	(27.32)
Número de	média	5.93	4.09	3.80	6.08	28.125	25.46154
visitas (vc)	(d.p.)	(5.83)	(3.64)	(3.08)	(4.45)	(12.29)	(12.63)
Duração média	média	30.52	33.19	29.90	21.66	2.799583	3.17
das visitas (vm)	(d.p.)	(24.83)	(30.07)	(21.67)	(21.87)	(0.89)	(1.67)
Duração total	média	79.72	77.79	71.83	81.23	76.49833	70.23346
das visitas (vt)	(d.p.)	(25.00)	(24.13)	(35.35)	(34.81)	(37.80)	(30.05)
Duração total do	média	84.82	82.57	76.27	88.34	93.34417	81.54
Teste (dur)	(d.p.)	(28.15)	(25.49)	(36.88)	(34.31)	(42.91)	(32.80)

Fonte: Elaborado pelo autor

5. RESULTADOS E ANÁLISES

Nesta seção, apresentam-se de forma resumida o que era esperado, as estatísticas e uma análise, que inclui possíveis explicações para os resultados encontrados. A mesma lógica é aplicada na análise dos três testes. Para os Teste 1 e Teste 2 consideram-se os resultados das perguntas de compreensão e as métricas geradas pelo *Eye-tracker*. O Teste 3 não foi acompanhado de perguntas, então conta apenas com a análise do movimento ocular.

5.1 Resultados e Análises do Teste 1: Legibilidade sintática e léxica

A literatura de relatórios populares supõe que elaborar textos mais legíveis para o cidadão não especialista, aumenta sua compreensão do que se relata. O primeiro teste manipula, de forma isolada, a legibilidade de um relato de desempenho governamental em relação a complexidade sintática e sofisticação léxica. A Tabela 9 abaixo sintetiza as estatísticas geradas com as respostas das perguntas que medem a compreensão (pergunta 1 como proxy da atenção e pergunta 2 como proxy da interpretação) do texto apresentado no primeiro teste.

Espera-se que a taxa de acerto das perguntas 1 e 2 (t1q1 e t1q2) seja maior entre os participantes que acessaram o teste com a versão mais legível do texto (T1A). De fato, na análise de diferenças de médias simples, os participantes que tiveram acesso ao texto mais fácil apresentaram uma taxa de acerto média 30% maior na questão 1 (proxy de atenção). Porém, ao contrário do que se esperava, na questão 2 (*proxy* para compreensão), a taxa foi 10% menor.

Teste T de Student foi realizado para testar a Hipótese de que a diferença entre as médias era diferente de zero e maior entre os participantes que fizeram o teste com maior legibilidade ($A > B$). O teste T é o mais indicado pois as comparações são feitas entre duas amostras independentes (lembrando que o desenho do experimento é entre sujeitos). A pergunta de atenção (t1q1) deu uma diferença significativa e positiva, como esperado, com nível de significância a 1%. Já a pergunta de compreensão (t1q2) deu uma diferença negativa e significativa, ao contrário do que se esperava.

Com base na literatura revisada, também é possível supor que a taxa de acerto da pergunta de compreensão será maior entre os respondentes que acertaram a pergunta 1, independentemente da versão do teste a qual tiveram acesso. Entretanto, diferentemente do que se esperava, a média de acerto entre esses dois grupos foi praticamente a mesma (próximo a

0.6 em ambos os grupos). A influência da manipulação surte um efeito contrário, averiguado pelo Teste T, a um nível de significância de 10%. Ou seja, entre os participantes que acertaram a pergunta 1 de atenção, o texto mais legível afetou negativamente a compreensão. As médias, desvio padrão e resultados desse teste podem ser observados nas últimas linhas da tabela.

Tabela 9 - Teste de média (Teste T) para perguntas de Compreensão do Teste 1

Perguntas de Compreensão por teste	n	Media	D.p	Teste T - Pro (T>t)
Pergunta 1 (t1q1) Atenção A > B				
T1A	24	0.875	0.338	0.0093 ***
T1B	26	0.577	0.503	
Pergunta 2 (t1q2) Interpretação A > B				
T1A	24	0.500	0.510	0.9139
T1B	26	0.692	0.470	(-)0.0862 *
Pergunta 2 (t1q2) Interpretação A > B sendo t1q1=1				
T1A	21	0.476	0,512	0.9350
T1B	15	0.733	0.457	(-) 0.0650 *

Fonte: elaborado pelo autor

Nota: * sig a 10%, ** sig a 5%, *** sig a 1%

Nota 2: Dados com distribuição normal, averiguada pelo teste de Shapiro-Wilk.

Acertar a questão 1 dependia da tarefa cognitiva de memória de curto prazo, que incluso foi facilitada pelo experimento, pois a resposta estava nas últimas linhas do texto. Mesmo assim, a legibilidade afetou os resultados. O resultado da segunda pergunta, com foco na capacidade dos participantes de interpretar os relatórios de governos, mostra que um texto mais legível pode não aumentar a compreensão de pessoas que não tem conhecimentos de contabilidade pública.

Da literatura, sabe-se que o processo de leitura e compreensão de textos inclui diferentes atividades cognitivas, que envolvem tarefas mais simples, como as de relembrar (encontrado na literatura como *recall*) de curto prazo de informações explícitas (propostas pela pergunta 1). Até às mais complexas, como de interpretação e extração de significados, que possibilitam o entendimento de informações implícitas (propostas pela pergunta 2). Nesse teste, observa-se que a legibilidade (em relação a complexidade léxica e sintática) afetou as atividades cognitivas mais simples, mas não ajudou a melhorar o entendimento de informações implícitas.

Comparando os respondentes que acertaram a pergunta de atenção (t1q1=1) e os que erraram, é possível depreender que mesmo fazendo a leitura com certo nível de atenção, o entendimento de informações implícitas dos relatórios de governos pode não ser uma tarefa

simples. Em realidade, o texto mais complexo parece ter surtido um efeito contrário entre os que acertaram a pergunta de atenção. Na última linha e coluna da tabela 9, consta que, a um nível de significância de 10%, a diferença entre as médias de A e B foi negativa. Enfim, é possível que as frases e palavras mais longas tenham afetado a compreensão de forma positiva, pois ao encontrar essa dificuldade no texto, o participante precisa realizar mais esforço cognitivo, começa uma atividade cognitiva mais complexa e acaba assimilando aquilo que se deixou implícito.

O *Eye-tracker* foi usado para fornecer mais evidências sobre as atividades cognitivas realizadas pelos participantes durante os testes e sobre seu nível de compreensão. Esperava-se que todas as métricas de fixação e de visitas às AOIs fossem menores na versão do texto que a legibilidade é maior (T1A). A AOI de interesse da análise é a que cobre todo o texto, a ela se designou a sigla a7. A normalidade da distribuição dos dados foi averiguada pelo teste de Shapiro Wilk. A única métrica com distribuição normal foi a do tempo de fixação total, com a soma de todas as fixações realizadas (ft_t1ab_a7). A tabela 10 abaixo apresenta média, desvio padrão e um resumo dos testes estatísticos realizados.

Normalmente se assume que, quando um texto é mais difícil, os leitores fazem fixações mais longas, sacadas mais curtas e mais regressões (Rayner, 2006). Esses movimentos oculares são refletidos nas métricas. Assim, espera-se que as métricas de quem viu a versão mais legível do texto sejam todas menores do que as métricas de quem viu a versão menos legível. Entretanto, os testes estatísticos apontam que não existe uma diferença de médias significativa entre os grupos. Nem mesmo o tempo total que o participante leva no experimento até dar o clique no mouse para trocar de tela (dur_t1ab) foi sensível.

É possível que a diferença de legibilidade (resultado do comprimento das palavras e frases) não tenha alterado a dificuldade do texto. Apesar de o índice de legibilidade calculado (ARI) ser o dobro no texto menos legível, as alterações são de fato superficiais. Por exemplo, retira-se o ponto, e duas frases passam a ser uma mais longa. Uma palavra mais curta, como “veio de” foi substituída por “advindo de”. Outro fato, é que a passagem utilizada, por uma limitação de tempo de colaboração do participante para realização do experimento, pode ter sido curta. Esse resultado vai no sentido das críticas realizadas às formulas de legibilidade, que afirmam que os índices não são realmente úteis para prever compreensão porque não são capazes de medir a complexidade da narrativa. Mesmo havendo uma grande diferença entre os índices calculados (Tabela 5), os textos apresentaram uma complexidade muito similar, pois

ainda tratam do mesmo assunto, carregam a mesma narrativa.

Tabela 10 - Testes de média para métricas do *Eye-tracker* geradas no Teste 1

Métricas geradas pelo <i>Eye-tracker</i> para AOI a7 (texto completo)		n	Media	D.p	Prob > z ; Prob B > A
Tempo Total de fixação	T1A	24	71.35	23.62	Ns
	T1B	26	67.86	23.66	
Média das fixações	T1A	24	0.386	0.61	Ns
	T1B	26	0.373	0.05	
Número de fixações	T1A	24	183.5	64.93	Ns
	T1B	26	181.61	58.49	
Tempo total da visita	T1A	24	79.72	25.00	Ns
	T1B	26	77.79	24.13	
Média das visitas	T1A	24	30.52	24.83	Ns
	T1B	26	33.19	30.07	
Número de visitas	T1	24	5.93	5.83	Ns
	T1B	26	4.09	3.64	
Tempo de leitura total.	T1A	24	84.82	28.15	Ns
	T1B	26	82.57	25.49	

Fonte: elaborado pelo autor

Nota: * sig a 10%, ** sig a 5%, *** sig a 1%

Nota 2: Com exceção da métrica ft_t1ab_a7 apresentada na primeira linha, os dados não apresentam distribuição normal (averiguada pelo teste de Shapiro-Wilk). Por isso, para testar a hipótese nula usamos o teste de Mann-Whitney mais (análise não-paramétrica).

Nota 3: Para a métrica citada, que consta na primeira linha da tabela, foi usado o Teste T.

Nota 4: O primeiro valor da última coluna (Prob > |z|) é a probabilidade resultado do teste da hipótese nula de que as médias são iguais. O segundo valor é a probabilidade de que a métrica seja maior entre os participantes do grupo A (Prob A > B).

Soma-se o fato de que a elaboração dos textos usados no Teste 1, dada a validade interna do experimento, considerou que as informações contidas em ambos os textos trechos deveriam ser iguais. Também o tom da narrativa, sem um viés negativo ou positivo do desempenho do governo (como é proposto pela literatura de RPs – veja seção 4.7.1 para mais detalhes). Enfim, os textos eram, essencialmente, muito similares. Essas são possíveis explicações para a ausência de diferenças entre os movimentos oculares dos grupos A e B.

5.2 Resultados e Análises do Teste 2: Legibilidade pela presença de jargões

Outra diretriz importante dentro da literatura de relatórios populares é que os jargões da contabilidade pública deveriam ser evitados. Supõe-se que o Relatório Popular poderia ser

melhor compreendido se os termos técnicos e específicos fossem substituídos por palavras mais “amigáveis” (Sharp *et al.*, 1998). Assim, no segundo teste, foi manipulada a presença de jargões em um relato de desempenho governamental.

Com base na literatura de relatórios populares, espera-se que a taxa de acerto das perguntas de atenção e de interpretação (t2q1 e t2q2) seja maior entre os participantes que acessaram o teste com a versão do texto sem jargões (T2A). A média de acerto da pergunta de atenção (t2q1) entre os respondentes que viram a versão sem jargões foi 13% maior, enquanto a média da pergunta de interpretação (t2q2) foi 3% menor. O teste de Mann-Whitney foi realizado e testou a hipótese nula de que não há diferença entre as médias, porém os resultados não foram significantes, como mostra a Tabela 11.

Tabela 11 - Testes de média para Compreensão do Teste 2

Perguntas de Compreensão por teste	n	Media	D.p.	Prob > z ; Prob A > B
Pergunta 1 (t2q1) Atenção A > B				
T2A	26	0.961	0.196	ns
T2B	24	0.833	0.380	
Pergunta 2 (t2q2) Interpretação A > B				
T2A	26	0.884	0.325	ns
T2B	24	0.916	0.282	
Pergunta 2 (t2q2) Interpretação A > B	Sendo t2q1=1			
T2A	25	0.92	0.276	ns
T2B	20	0.95	0.220	

Fonte: elaborado pelo autor

Nota: * sig a 10%, ** sig a 5%, *** sig a 1%

Nota 2: Dados não apresentaram distribuição normal (averiguada pelo teste de Shapiro-Wilk). Para testar a hipótese nula usamos o teste de Mann-Whitney.

Nota 3: O primeiro valor da última coluna é a probabilidade resultado do teste da hipótese nula de que as médias são iguais. O segundo valor após o ‘;’ (Prob A > B) é a probabilidade de que a métrica seja maior entre os que fizeram o teste T2A.

Ao ler, o indivíduo poderia encontrar dificuldade nos processos de reconhecimento e decodificação do significado dos jargões. Diferentemente do teste 1, que analisou a dificuldade geral encontrada na narrativa do texto, a proposta desse teste está focada na análise do processamento cognitivo local, de termos específicos. As AOIs foram configuradas para que se possa inferir, de forma mais precisa, sobre o movimento ocular ao ler cada um dos jargões. Além da AOI que cobre todo o texto (a5), destacam-se outras 3 (a1, a2 e a3), que contém os

jargões na versão tradicional do relato (T2B e, na versão “traduzida” (T2A), cobrindo os termos que os substituíram.

Espera-se que, no Teste 2B, as métricas de fixação e visita sejam maiores. Para averiguar se essa diferença existe e é significativa, foram realizados os testes T de Stuart (para as métricas com distribuição normal) e Mann-Whitney (não-paramétrico). Os jargões utilizados e suas respectivas traduções são apresentados na tabela 12 a seguir para contextualizar a interpretação dos dados. Os resultados são apresentados, de forma resumida, na tabela 13 logo abaixo.

Tabela 12 - Descrição das AOIs configuradas para análise do Teste 2 (efeitos da presença de jargões)

Sigla da AOI	Descrição	Termos "traduzidos" - T2A	Jargões da contabilidade - T2B
a1	Jargão 1	decisão do governo de não cobrar tributos	renúncia de receita aprovada pelo prefeito
a2	Jargão 2	os prejuízos futuros da previdência	passivo atuarial da previdência
a3	Jargão 3	receita arrecadada foi 27% menor do que	teve uma frustração de receita de 27%.
a5	texto	texto completo	texto completo

Fonte: elaborado pelo autor

De acordo com a tabela, não houve uma diferença significativa entre as métricas geradas nos testes T2A e T2B para as AOIs a1 e a3. Ou seja, aparentemente, durante a leitura, os jargões “frustração de receita” e “renúncia de receita” não foram percebidos como sendo mais difíceis a ponto de se ter que fazer mais fixações (e fixações mais longas) ou mais visitas (e visitas mais longas) para seu reconhecimento e decodificação. Lembrando que, na validação dos jargões realizada previamente com participação da própria população amostral, todos os termos “frustração de receita”, “renúncia de receita” e “passivo atuarial” foram percebidos como difíceis ao serem analisados isoladamente (mais detalhes na seção 4.6 e no apêndice)

Por outro lado, as estatísticas mostram que houve uma diferença significativa do movimento ocular dos respondentes que leram o trecho com jargão “passivo atuarial da previdência”, e sua versão traduzida “os prejuízos futuros da previdência”. A média dos tempos de fixação (sig. 10%), o número de fixações (sig. 5%) e número de visitas (sig. 5%) foram maiores. De acordo com a literatura em movimentos oculares, essa é uma evidência de que houve uma dificuldade de leitura local, de forma que o indivíduo teve de colocar essa área (a AOI a2) no centro do seu campo de visão (fóvea) mais vezes.

Analisando as métricas para o texto completo (a5), observa-se que o tempo total de fixação (ft_t2ab, sig. 10%), o número de fixações (fc_t2ab, sig. 10%), o número de visitas

(vc_t2ab, sig. 5%) e o tempo total de leitura (dur_t2ab, sig. 10%) foram maiores para o teste com jargões (T2B). A duração média das visitas se mostrou diferente, porém menor no teste com jargões (sig. 10%). São evidências clássicas de dificuldade de compreensão de um texto. O fato de as visitas médias serem menores se dá pois, ao fazer regressões, é normal que as visitas sejam mais rápidas (Rayner, 1998). Tipicamente, é a primeira visita/fixação que demora mais. Depois, quando os olhos voltam a visitar determinada palavra ou sentença, o tempo que ele fica ali focado é menor (Rayner, 1998).

O comportamento ocular indica que a presença dos jargões alterou a dificuldade geral do texto. Apesar de não se registrarem medidas significativamente diferentes em dois dos três jargões, as evidências apontam, de forma clara, que os indivíduos encontraram mais dificuldade no processamento da narrativa no texto com jargões. Com base nesses resultados, o mais provável é que as perguntas de compreensão elaboradas tenham falhado no seu objetivo de medir o nível de compreensão do texto. É necessário reconhecer que medir a compreensão de um texto, para identificar se o indivíduo pode extrair mensagens que não estavam explícitas, também não foi uma tarefa fácil.

O fato de várias métricas sobre o texto completo apresentarem uma diferença significativa é interessante pois fica claro que, apesar de não ter dificuldade de processamento local sobre as palavras monitoradas (jargões vs substitutos), o indivíduo claramente tem mais dificuldade para processar o texto como um todo. Supõe-se que os jargões não necessariamente atrapalham a fluência (fluxo) da leitura, mas geram confusão, fazendo com que o indivíduo faça mais regressões (voltando a trechos anteriores que já haviam sido lidos). O indivíduo lê, porém não forma o significado completo da narrativa, a leitura segue, mas fica ali um vazio de significado, que pode ser chamado de *blanks*. Devido aos vazios, há a necessidade de buscar sentido no texto fazendo mais visitas e o que leva a um tempo total de fixação no texto maior, reflexo de uma maior dificuldade de processamento das informações.

Tabela 13 - Testes de média para métricas do *Eye-tracker* no Teste 2

Métricas (texto completo)	Prob > z	Métricas (jargão 2)	Prob > z
T2A n =26 / T2B n = 24	Prob B > A	T2A n =26 / T2B n = 24	Prob B > A
Tempo Total de fixação	0.0999*	Tempo Total de fixação	0.0730* 0.673
Média das fixações	ns	Média das fixações	Ns
Número de fixações	0.0818*	Número de fixações	0.0256** 0.712
Tempo total da visita	ns	Tempo total da visita	Ns
Média das visitas	0.0709* 0.351	Média das visitas	ns
Número de visitas	0.0299** 0.678	Número de visitas	0.0223** 0.711
Duração do teste (tempo de leitura total)	0.0709* 0.649		
Métricas (jargão 1)		Métricas (jargão 3)	
T2A n =26 / T2B n = 24	Prob > z	T2A n =26 / T2B n = 24	Prob > z
Tempo Total de fixação	ns	Tempo Total de fixação	ns
Média das fixações	ns	Média das fixações	ns
Número de fixações	ns	Número de fixações	ns
Tempo total da visita	ns	Tempo total da visita	ns
Média das visitas	ns	Média das visitas	ns
Número de visitas	ns	Número de visitas	ns

Fonte: elaborado pelo autor

Nota: * sig a 10%, ** sig a 5%, *** sig a 1%

Nota 2: Para mostra com distribuição normal (averiguada pelo teste de Shapiro-Wilk) foi realizado Teste T, caso contrário, teste de Mann-Whitney.

Nota 3: Prob > |z| é a probabilidade da hipótese nula de que as médias são iguais. Prob B > A é a probabilidade de que a métrica seja maior entre os participantes do grupo B

5.3 Resultados e Análises do Teste 3: Legibilidade de tabelas com jargões

Como foi explicado anteriormente, os Relatórios Populares deveriam usar menos termos técnicos, que tem significados muito específicos da contabilidade e das finanças públicas. Sobre os efeitos da presença de jargões em tabelas, não foram encontrados estudos dentro da literatura de legibilidade ou de movimento ocular durante a leitura que pudessem prever como seria o comportamento ocular dos indivíduos. Assim, a análise dos resultados desse teste parte de um princípio mais exploratório, sem um pressuposto principal. Assim mesmo, se os achados encontrados nos estudos que usam trechos de textos por extenso se aplicassem nesta análise, seria razoável esperar que as métricas geradas sobre os termos menos familiares (jargões contábeis dos relatórios tradicionais – teste T3B) fossem maiores do que as métricas geradas para os termos substitutos (simplificados). Ou seja, com base na literatura sobre leitura, espera-se que, ao deparar-se com termos mais difíceis, os participantes tenham que fazer mais fixações e visitas (e mais longas) para concluir o reconhecimento e decodificação dos significados. Sem embargo, na comparação das AOIs que cobrem todo o ativo, e todo o passivo, não foram encontradas diferenças significativas que corroborem esta suposição.

A tabela 14 a seguir apresenta os resultados dos testes de média realizados. Nenhuma das métricas se mostrou sensível à manipulação. Ou seja, não se constatou uma diferença significativa no padrão de movimento ocular sobre a área que cobre o ativo, e sobre a área que cobre o passivo.

Tabela 14 – Comparação do lado esquerdo (ativos) e direito (passivos) das Tabelas no Teste 3

Métricas AOI a4 (Tabela E - ativos)		Métricas AOI a3 (Tabela D - passivos)	
T3A n=24	Prob > z ; Prob B > A	T3A n=24	Prob > z ; Prob B > A
T3B n = 26		T3B n = 26	
Tempo Total de fixação	ns	Tempo Total de fixação	ns
Média das fixações	ns	Média das fixações	ns
Número de fixações	ns	Número de fixações	ns
Tempo total da visita	ns	Tempo total da visita	ns
Média das visitas	ns	Média das visitas	ns
Número de visitas	ns	Número de visitas	ns
Tempo total de leitura	ns		

Fonte: elaborado pelo autor.

No caso do experimento 3, o mais provável é que outros fatores tenham interferido no comportamento ocular dos indivíduos. A pergunta deixada logo abaixo da tabela, “Qual a situação financeira da prefeitura?” deve ter influenciado o comportamento visual. Assim, é possível supor que, mesmo diante de uma tabela com palavras difíceis, os participantes de ambos os grupos se encontravam em uma tarefa cognitiva similar, de formular uma resposta para a pergunta. Essa situação pode explicar a semelhança entre as métricas de A e B na comparação entre ativo e passivo.

Esse teste foi desenhado com o objetivo de se fazer uma análise exploratória, imaginando que os termos simplificados pudessem alterar o padrão de movimento ocular desempenhado. Para realizar essa análise foram criadas outras variáveis, apresentadas na tabela 15. Número / texto calcula o quanto os participantes olham para os números em relação ao texto. Essa foi a única dentre as novas variáveis criadas que se mostrou significativa.

Tabela 15 – Testes de Média (Mann-Whitney) das variáveis criadas para análise do Teste 3

Métricas	Média (D.p.)	Média (D.p.)	Prob > z
Número / Texto	A n = 24	B n = 26	
Número de fixações	1.38 (0.58)	0.98 (0.56)	0.0029***
Tempo Total de fixação	1.33 (0.51)	1.10 (0.66)	0.0570**
Visita Média	1.19 (0.34)	1.08 (0.40)	ns
Ativo / Passivo	A n = 24	B n = 26	
Número de fixações	1.38 (0.97)	1.18 (0.39)	ns
Tempo Total de fixação	1.36 (0.84)	1.26 (0.50)	ns
Visita Média	1.33 (0.72)	1.27 (0.45)	ns
Ano atual / Ano anterior	A n = 24	B n = 26	
Número de fixações	2.102 (1.15)	2.22 (1.74)	ns
Tempo Total de fixação	2.16 (1.26)	2.27 (2.03)	ns
Visita Média	1.48 (0.74)	1.46 (0.95)	ns

Fonte: Elaborado pelo autor.

Os participantes que viram a versão A do teste (simplificada), colocaram mais atenção visual nos números, em relação ao conteúdo textual da tabela. Esse resultado é interessante. Significa que o participante não tem que fazer tanto trabalho cognitivo para ler as palavras, uma vez que já as compreendeu, o foco é colocado nos números da tabela.

As variáveis Ativo / Passivo e Ano Atual / Ano anterior foram criadas para identificar se, ao ler os relatórios simplificados, o participante apresentaria um padrão de movimento ocular com algum viés para essas classes de informação. Os testes de média indicam que não

houve uma diferença significativa entre o grupo que viu a versão A e a versão B. Portanto, é possível afirmar que a simplificação dos relatórios não levaria a um viés sistemático de atenção visual, nem para o tipo de conta, nem para o período do exercício fiscal.

6. IMPLICAÇÕES

O conjunto de evidências aqui levantado pode servir de base para os formuladores de políticas públicas, contadores públicos que elaboram os relatórios e seus parceiros na elaboração de Relatórios Populares melhores. Essa entrega colabora no sentido de diminuir a ilusão de transparência atual. Apesar da abertura de dados, com um grande volume de dados sendo disponibilizado nos portais web, pouco tem sido feito para aumentar a transparência efetiva em relação ao desempenho de governos.

Os resultados apresentados mostram que a análise de compreensão deve considerar as diferentes tarefas cognitivas que os indivíduos desempenham ao ler relatórios sobre desempenho de governos. Primeiro, os resultados indicam caminhos para a composição de relatórios populares, especificamente na substituição de jargões da contabilidade por palavras mais familiares, acompanhadas de contexto ou de exemplos que ajudem os leitores a fazer sentido dos conteúdos apresentados.

O RP já foi colocado como uma versão traduzida de um relatório completo (Paolo, Silvana & Valerio, 2016). Mas, por vezes, tal tradução se restringe à um efeito visual do relatório. A iniciativa de Relatório Popular emitido pela secretaria da Fazenda do Estado da Bahia, já citado nesta pesquisa, faz uso de gráficos, traz ilustrações e muitas cores, porém continua carregado de jargões contábeis. As cores, gráficos, frases mais curtas e fotos ajudam, mas não resolvem o problema da compreensão para não especialistas. É preciso, além de simplificar as frases e as palavras, substituir os jargões e adicionar informações de contexto, mesmo que isso necessariamente leve a perdas no que tange a precisão ou acuracidade das informações.

O desafio do elaborador do RP é calibrar o mesmo código de significados do relatório para se conectar com significados latentes utilizados na linguagem fluída e instável da sociedade. Os termos ou as palavras podem ser vistos como conjuntos de símbolos que remetem a um determinado significado. Substituir as palavras seria trocar esses símbolos por outros com mesmo significado. Contudo, algumas palavras não possuem substitutas diretas.

Na presente pesquisa, os trechos de relatórios populares para os três testes foram gerados com e sem jargões. Percebeu-se que a missão de “traduzir” um relatório financeiro de uma entidade pública substituindo os termos técnicos por palavras menos estranhas não é uma tarefa simples. O assunto do qual o relatório trata é técnico, e expressá-lo usando palavras

menos técnicas, buscando transmitir os mesmos significados é mesmo difícil.

O desafio é encontrar no vocabulário popular o conjunto de palavras que tem o mesmo significado desejado. Se não há, o significado deve ser desenvolvido. Se o significado que precisa ser debatido – passivo atuarial da previdência, por exemplo – não existe no vocabulário popular, não pertence a lógica institucional social, o tal trabalho de tradução será ainda mais difícil. Uma solução é aceitar a perda de parte do significado que se pretende transmitir, com o uso de um termo que expressa algo mais específico ou mais amplo do que aquele que o jargão intencionava retratar. Entretanto, se esse termo entregar significado, mesmo que não seja o mais preciso deles, ao menos entregará uma parte da informação.

Tal perda de informação seria uma alternativa ao uso de jargões técnicos e tabelas que são precisas, porém com baixa capacidade de entrega de significado ao usuário não especialista. Os responsáveis pela elaboração dos relatórios talvez não estejam atentos à esta questão.

O fato é que certas palavras, aqui tratadas como ‘jargões’, geram o que aqui - indutivamente - se chamou de ‘*blanks*’ (vazios de significado). Essa é uma questão a ser analisada detalhadamente, mas supõe-se que tais *blanks* de significado dificultariam ou impediriam o entendimento do leitor. Jargão, como tratado pela literatura de legibilidade, é um termo não-familiar, de significado complexo ou dubio. Se dentre os indivíduos sem conhecimento especializado existem vários grupos, cada um tem conhecimento de certos termos e não de outros, será quase impossível reduzir o número de jargões a zero.

Segundo ponto a ser considerado como implicações desses resultados, é que diferentes tarefas cognitivas podem ser mais ou menos facilitadas pela simplificação da linguagem no RP. A simplificação do texto (melhora da legibilidade) pela retirada de jargões e outras formas não garante um bom desempenho em todas tarefas cognitivas envolvidas na compreensão do relatório. Tarefas que usam memória de trabalho podem ser facilitadas, mas outras mais complexas de interpretação talvez não. Ao considerar as diferentes tarefas cognitivas, o uso combinado de contextualização dos eventos, associação com outros eventos e fatos que compõem a história do período de desempenho em análise ou história passada, assim como uma maior associação com a realidade do usuário, poderia facilitar o processo de interpretação.

Entre as tarefas cognitivas estão: a recuperação de significados pré-existentes para uma palavra do texto que desvende uma ação consciente cognitiva de associação por similaridade / dissimilaridade cujo resultado seria o transporte de significado registrado na memória à mesa de processamento corrente (recorrente na literatura como ‘*working memory*’). Essa associação

não necessariamente seria como o significado exato daquele termo (desejado por quem elaborou o relatório), mas aquele conceito mais próximo e inconscientemente associado ao termo.

Esses conceitos não são trabalhados de forma isolada, mas emergem de uma concatenação lógica dos significados recuperados na construção de uma narrativa que representa o que foi lido. Esse seria um caminho a ser explorado nos testes empíricos de relatórios populares, ou seja, analisar as diferentes tarefas cognitivas presentes e necessárias ao usuário que é exposto ao relatório.

Por fim, em geral, quando se trata de relatório popular, assume-se um relatório escrito, com tabelas, gráficos e fotos. Mas o conteúdo sobre desempenho pode ser transferido por outras mídias. A divulgação de informações por vídeos curtos em mídias sociais pode ter melhores resultados, por combinar melhor outros elementos de linguagem, favorecendo a compreensão da mensagem. Ainda, em todo relatório existe uma lógica de narrativa implícita, que surge na interação do leitor com o relatório. A lógica de exposição dos argumentos no relatório poderia ser potencializada, melhor conduzida com uso de outras mídias.

7. CONCLUSÕES

Nas perguntas de compreensão dos testes 1 e 2, foram propostos dois desafios cognitivos. No primeiro, o respondente deveria simplesmente recuperar uma informação que estava explícita no texto lido, o que está associada à memória de trabalho. Nesse caso, o texto mais legível melhorou os resultados. Ou seja, o texto mais legível facilita a captação e a memorização de informações explícitas. Quando o indivíduo está ocupado tentando silenciosamente fazer sentido de um texto mal escrito, sua memorização é prejudicada.

Já no outro desafio, proposto pela segunda pergunta de compreensão, o respondente teria que interpretar o texto, para entender uma condição que estava implícita. Nesse caso, a legibilidade não melhora os resultados. Parece que essa tarefa depende mais do leitor, do que do texto em si, o que é condizente com a literatura de legibilidade. Em suma, a manipulação da complexidade sintática e sofisticação léxica (medida pelo comprimento de frases e palavras) melhora a memorização, mas não necessariamente aumenta a compreensão, pois não ajuda o leitor a interpretar os textos. Assim, a facilitação do texto em relação a complexidade sintática e léxica melhora a capacidade de recuperação de informações, mas talvez não seja suficiente para a tarefa de interpretação.

O teste 1 manipula a legibilidade em quanto a complexidade léxica e sofisticação sintática. Ou seja, o texto mais legível apresenta a mesma narrativa (carregando as mesmas informações) através de palavras e frases mais curtas. Apesar da diferença de legibilidade, calculada pelo índice ARI, o padrão de movimento ocular (refletido nas métricas captadas) não parece ter mudado. O resultado vai ao encontro das críticas às fórmulas de legibilidade. Outras pesquisas poderiam ser realizadas, considerando também outros índices, para ver se outra fórmula reflete melhor a compreensão de relatórios contábeis ou financeiros do setor público.

O teste 2 trata da presença de jargões em textos, uma condição muito discutida na literatura de legibilidade com o nome de palavras não-familiars. No teste aplicado, os jargões não afetaram a fluência da leitura, ao contrário do que se esperava. Entretanto, provocaram uma alteração no comportamento ocular no texto como um todo. O fato de várias métricas terem apresentado uma diferença significativa indica que o leitor teve que fazer mais regressões dentro do texto (revisitando partes que já haviam sido focadas). Esse resultado indica que, apesar de não haver evidências de dificuldade de processamento cognitivo local em dois dos três jargões observados, o padrão de movimento ocular sinaliza uma dificuldade generalizada

de leitura, o que pode ser entendido como mais trabalho cognitivo para compreensão. Em outras palavras, o efeito dos jargões não necessariamente diz respeito ao aumento do tempo de leitura por paradas no texto, mas pode gerar a criação de vazios de significado no processo de leitura, acumulando o que chamamos de vazio de significado. As regressões dentro do texto seriam como se o respondente tentasse preencher esses vazios.

No teste 3, em que se observam os efeitos dos jargões em tabelas, as tabelas tinham áreas de conteúdo informacional distintas pois apresentam uma estrutura de balanço patrimonial. A presente pesquisa pode comparar conteúdos como passivo e ativo, números e texto, valores atuais e anteriores, mas outras comparações seriam possíveis, como por exemplo valores mais agregados e valores menos agregados. O aumento da legibilidade do texto na tabela por substituição de jargões não alterou a região de atenção quanto a ativo e passivo e valores atuais e anteriores. Contudo, aumentou a procura por significados com os olhos focados nos números, em comparação aos textos. Esse achado indica que de fato, os jargões contábeis aumentam a dificuldade que jovens indivíduos têm para interpretar os resultados. Pois no caso em que os jargões foram substituídos por termos simples, o foco do usuário se concentra nos números.

Cabe ressaltar que não necessariamente os indivíduos extraíram significados e interpretações claras. Como no caso do Teste 2, que apesar de não demonstrar dificuldade no processamento local dos jargões, foi possível detectar uma dificuldade geral. Sugere-se que sejam realizados outros estudos, buscando explorar, por meio de uma abordagem qualitativa, qual seria a qualidade dos significados extraídos. É possível supor que, mesmo os termos simplificados, tenham sido de difícil entendimento. Apesar de serem apresentadas palavras menos técnicas, em essência, o significado por trás das “traduções” segue sendo o mesmo. Em outras palavras, por mais que a tabela deixe de trazer complexidade com jargões não conhecidos, a lógica da tabela ainda contempla conceitos contábeis, da comunicação de desempenho financeiro. O que, para muitas pessoas sem conhecimentos ou experiência no assunto continua sendo um tema desafiador e incompreensível.

Em suma, os testes levantam evidências de que um texto mais legível em relação a dificuldade sintática e léxica pode melhorar o resultado de tarefas cognitivas como a memorização de informações relevantes e a recuperação de conceitos vistos no texto. Além disso, a presença de jargões piora a compreensão, afetando o processamento cognitivo geral (do texto) e local (dos termos).

A pesquisa explorou a frente de ação da simplificação do texto para aumento da compreensão dos relatórios populares, porém deve-se ter em mente que essa frente de atuação também tem um limite, pois parte do resultado (compreensão final) depende do leitor, das suas capacidades cognitivas. Portanto, parte da solução está no melhoramento do relatório, enquanto a outra está no desenvolvimento cognitivo dos indivíduos, o que está associado a melhores níveis de educação.

REFERÊNCIAS

Clifton Jr, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In *Eye Movements* (pp. 341-371). Elsevier.

AGA. 2012. E-Reporting. **Corporate Partner Advisory Research Series**, Report No. 32. Disponível em: <<https://www.agacgfm.org/getattachment/Resources/Online-Library/Research-Reports/EReportingJuly2012.pdf.aspx>>. Acessado em: janeiro de 2018.

Allison, G.S. (1995). Gafo-as Popular Reporting Awards Program: Revised and Improved. **Government Finance Review**, 11 (3), p. 36–37.

Barbera, C.; Borgonovi, E., & Steccolini, I. (2016). Popular Reporting and Public Governance: The Case of ‘Bilancio in Arancio’ in Milan Municipality. **Governance and Performance in Public and Non-Profit Organizations** (Studies in Public and Non-Profit Governance), 5, p. 3– 30.

Brown, A. D., Colville, I., & Pye, A. (2015). Making sense of sensemaking in organization studies. **Organization Studies**, 36(2), 265-277.

Brusca, I., & Montesinos, V. (2006). Are citizens significant users of government financial information? **Public Money and Management**, 26(4), p. 205-209.

Campos, A.M. (1990). Accountability: quando poderemos traduzi-la para o português? **Revista de Administração Pública**, 24 (2), p. 30-50.

Carpenter, F.H., & Sharp, F.C. (1992). **Popular reporting: local government financial reports to the citizenry**. Governmental Accounting Standards Board, Norwalk.

Senter, R. J., & Smith, E. A. (1967). *Automated readability index*. CINCINNATI UNIV OH.

Chall J.S. (1958). **Readability: An Appraisal of Research and Application**. Ohio State University Press, Columbus.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). **Experimental methods: Between-subject and within-subject design**. *Journal of Economic Behavior & Organization*, 81(1), 1-8.

Ceneviva, R., Farah, M. F. S. (2012). Avaliação, informação e responsabilização no setor público, **Revista de Administração Pública**, 46 (4), p. 993–1016.

Cohen, S., & Karatzimas, S. (2015). Tracing the future of reporting in the public sector: introducing integrated popular reporting. *International Journal of Public Sector Management*, 28(6), 449-460.

- Cohen, S.; Mamakou, X.; Karatzimas, S. (2016). IT enhanced popular reports: Is it the missing link in triggering citizens' interest? **Athens University of Economics and Business**.
- Colville, I., Brown, A. D., & Pye, A. (2012). Simplexity: Sensemaking, organizing and storytelling for our time. **Human Relations**, 65(1), 5-15.
- Dean, M. M. (2002). The role of GASB 34 in the citizen-government accountability relationship. **State and Local Government Review**, 34(1), p. 51-53.
- DuBay, W. H. (2004). **The Principles of Readability**. California: Impact Information, 73 p.
- Duchowski, A. T. (2007). Eye tracking methodology. *Theory and practice*, 328(614), 2-3.
- Eckersley, P., Ferry, L., & Zakaria, Z. (2014). A “panoptical” or “synoptical” approach to monitoring performance? Local public services in England and the widening accountability gap. **Critical Perspectives on Accounting**, 25(6), p. 529–538.
- Foucault M. Discipline and Punish. (A. Sheridan, Trans.) London: Allen Lane; 1977.
- Ganske, K., & Fisher, D. (2009). **Comprehension across the curriculum: Perspectives and practices K-12**. Guilford Press, 335 p.
- GFOA. (2018). Popular Annual Financial Reporting Award Program (PAFR Program). Disponível em: <<http://www.gfoa.org/pafr>>. Acessado em: março de 2018.
- Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, 77(1), 45-56.
- Haustein, E. & Lorson, P. C. (2017). **“I no longer trust the numbers” – A study on citizens' reflections on local government financial statements in the light of transparency and distrust**. Working Paper presented at the 16th Biennial CIGAR Conference – Comparative International Governmental Accounting Research Network Porto, Portugal.
- Heald, D. (2003). Fiscal transparency: Concepts, measurement and UK practice. **Public administration**, 81(4), p. 723-759.
- Hermann, K. R. (2011). Creating an Excellent Popular Financial Report. **Government Finance Review**, 27(5), p. 41-44.
- Hood, C. (2010). Accountability and Transparency: Siamese Twins, Matching Parts, Awkward Couple? **West European Politics**, 33(5), p. 989–1009.
- International Public Sector Accounting Standards Board (IPSASB). (2013). **Conceptual Framework for General Purpose Financial Reporting by Public Sector Entities**, International Public Sector Accounting Standards Board, New York, NY.
- Janan, D., Wray, D., & Pope, M. (2010). Paradigms in readability research. **International Journal of Arts and Sciences**, 3(17), 19-29.

Jordan, M. M., Yusuf, J. E., Berman, M., & Gilchrist, C. (2017). Popular Financial Reports as Fiscal Transparency Mechanisms: An Assessment Using the Fiscal Transparency Index for the Citizen User. **International Journal of Public Administration**, 40(8), 625-636.

Justice, J., & Dulger, C. (2009). Fiscal transparency and authentic citizen participation in public budgeting: The role of third-party intermediation. **Journal of Public Budgeting, Accounting & Financial Management**, 21(2), p. 254-288.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. **Scientific Studies of Reading**, 12(3), 281-300.

Kolstad, I., Wiig, A. (2009). Is Transparency the Key to Reducing Corruption in Resource-Rich Countries? **World Development**, 37(3), p. 521-532.

Lee, M. (2006). The history of municipal public reporting. **Intl Journal of Public Administration**, 29(4-6), p. 453-476.

Lindstedt C. & Naurin, D. (2010). Transparency is not Enough: Making Transparency Effective in Reducing Corruption. **International Political Science Review**, 31(3), p. 301-322.

Lourenço, É. A. G. D., Hayashi, M. C. P. I., & Almeida, M. A. D. (2009). Single-subject research design in theses and dissertations of the PPGEES/UFSCAR. **Revista Brasileira de Educação Especial**, 15(2), p. 319-336.

Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. **The Academy of Management Annals**, 8(1), p. 57-125.

Marsh, T.L; Montondon, L.G. (2005a) A Comparison of The Readability of Governmental Annual Financial Reports Popular Reports and Management Discussion and Analysis. **Journal of Accounting & Finance Research**, 13(3), p. 153-161.

Mathiesen T. (1997). The viewer society: Michel Foucault's 'Panopticon' revisited. **Theoretical Criminology**, 1, p. 215-34.

Montondon, L. G., & Lilley, T. (2005b). How Readable are Those Financial Reports? **The Journal of Government Financial Management**, 54(1), 52.

Observatório Social do Brasil. (2018). Disponível em: < <http://osbrasil.org.br/>>. Acesso em: fevereiro de 2018.

Oliveira, J. H. C. (2013). Neuromarketing, tomada de decisão e consumo sustentável entre jovens: um estudo sobre o processamento cognitivo duplo e seu impacto na brand equity por meio da neurociência visual. **Tese de Livre-Docência – Universidade de São Paulo**, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, 74 p; 31 cm.

Paolo, B., Silvana, S., & Valerio, B. (2016). The Popular Financial Reporting: Focus on Stakeholders—The first European Experience. **International Journal of Business and Management**, 11(11), p.115-125.

Rodríguez Bolívar, M. P., & Galera, A. N. (2016). The effect of changes in public sector accounting policies on administrative reforms addressed to citizens. **Administration & Society**, 48(1), p. 31-72.

Scholte, J. A. (2008). Reconstructing Contemporary Democracy. **Indiana Journal of Global Legal Studies**, 15(1), Article 13.

Sharp, F. C., Carpenter, F. H., & Sharp, R. F. (1998). Popular financial reports for citizens. *The CPA Journal*, 68(3), 34.

Stausberg, M., & Engler, S. (2013). **The Routledge handbook of research methods in the study of religion**. Routledge.

Sutcliffe, K. M. (2013). Sensemaking. *The Palgrave encyclopedia of strategic management*. Basingstoke: Palgrave Macmillan. Advance online publication.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. **Organization science**, 16(4), 409-421.

Wray, D., & Janan, D. (2013). Readability revisited? The implications of text complexity. **The Curriculum Journal**, 24(4), p.553-562.

Yusuf, J.E., & Jordan, M.M. 2012. Effective Popular Financial Reports: The Citizen Perspective. **Journal of Government Financial Management**, 61(4), p. 44-49.

Yusuf, J.E., Jordan, M.M., Neill, K.A., & Hackbart, M. (2013). For the people: Popular financial reporting practices of local governments. **Public Budgeting & Finance**, 33(1), p. 95-113.

APÊNDICE A – O protocolo do experimento

Os seguintes cuidados são tomados para manter a comparabilidade dos testes com o ET: fatores externos (ruído, temperatura ambiente e fontes de distração), e posicionamento dos participantes frente ao equipamento. O responsável pelo experimento, presente na sala de teste, opera o equipamento e o computador, e corrige o posicionamento do participante. O pesquisador passou o mínimo de tempo dentro do campo de visão do participante. Ressalta-se que em todas as etapas atuou-se de forma padrão afim de que a experiência fosse a mesma para todos participantes.

As condições físicas de uma sala totalmente dedicada para o experimento foram controladas pelos pesquisadores; entre elas, iluminação, temperatura e som ambiente. O teste não foi aplicado em momentos que haviam som fora do normal. A distância entre o aparelho *eye-tracker* e o participante também foi observada e mantida constante. A cadeira usada tinha regulagem de altura, que, como orienta o manual do equipamento, foi mantida entre 115 e 130 cm. Dessa forma o participante se mantinha dentro da área de 31° em relação ao centro do aparelho, como ilustra a Figura 2 abaixo.

Figura 2- Etapa 1: Preparação do participante e layout de coleta

Fonte: Oliveira (2013)

Num primeiro momento, o pesquisador interage com o participante orientando-o com relação a postura, distancia da tela e altura da cadeira. Depois, ele pede que o participante fixe os olhos no monitor, e acompanhe o círculo vermelho na tela. Esse procedimento sempre ocorre antes de cada coleta, esse procedimento se chama calibração e é importante para garantir o funcionamento adequado do equipamento. O pesquisador dá início ao experimento lendo em voz alta (sempre com um tom neutro): “O experimento vai começar. A partir desse momento, será você e a tela. Não vamos interagir durante o teste, é como se eu não estivesse aqui. Todas as instruções que você precisa estarão na tela. Para avançar, você deve dar um *click* com o mouse. Pronto? Começou.” Nesse momento o pesquisador inicia a gravação e o participante vê a primeira tela, que contém as instruções para os Testes 1 e 2. Após a tela de instrução, os participantes acessam o Teste 1 A, ou o Teste 1 B, que foram aleatoriamente designados. Ao terminar de ler o texto, o participante responde a duas perguntas de múltipla escolha na mesma tela do computador. Em seguida aparece o Teste 2, que da mesma forma é seguido de 2 questões. As figuras a seguir (Apêndice B) mostram as imagens originais a que os participantes foram expostos durante o experimento.

APÊNDICE B – Material dos testes 1 a 3

Figura 3 – Teste 1 A

A secretaria de infraestrutura faz a manutenção dos bens da prefeitura, além das obras e reformas públicas. Em 2017, o valor definido no orçamento para esta secretaria foi 10% maior, totalizando 33 milhões de reais. Parte das despesas foi destinada ao programa de melhoria da infraestrutura e mobilidade urbana. No programa, estão incluídos o recapeando de ruas e avenidas e reforma dos pontos de ônibus. Além de obras para melhorar a acessibilidade dos edifícios da prefeitura. De acordo com as demonstrações contábeis, o passivo aumentou 20% em relação a 2016. Esse aumento é consequência de um empréstimo obtido pela prefeitura. A verba foi usada para financiar as obras do programa de mobilidade urbana.

Além disso, a secretaria terminou de construir de três escolas cujas obras estavam paradas há cinco anos. Essa medida só foi possível usando receitas que não haviam sido previstas em orçamento. O dinheiro veio de doações recebidas de grandes empresários da região.

Figura 4 – Teste 1 B

A secretaria de infraestrutura é responsável pela manutenção dos bens da prefeitura, além de fazer as obras e reformas públicas. No ano de 2017, o valor que foi separado no orçamento para esta secretaria foi no montante de 33 milhões de reais, 10% a mais que 2016, sendo que parte das despesas foi destinada ao programa de melhoria da infraestrutura e mobilidade urbana. Entre as medidas do programa está incluso o recapeamento de ruas e avenidas, reforma dos pontos de ônibus e obras para melhorar a acessibilidade dos edifícios da prefeitura. De acordo com as demonstrações contábeis, o passivo aumentou 20% em relação a 2016, esse aumento é consequência de um empréstimo obtido pela prefeitura para financiar as obras do programa de mobilidade urbana citadas.

Além disso, a secretaria finalizou a construção de três escolas que tinham obras paradas há cinco anos refletindo em um aumento no ativo do município, isso só foi possível pois o governo contou com receitas que não haviam sido previstas em orçamento, advindas de doações recebidas de grandes empresários da região.

Figura 5 – Teste 2 A

No ano de 2017, a receita arrecadada foi 27% menor do que a receita que havia sido prevista. O planejamento de várias secretarias foi afetado, porém esse fato se refletiu, principalmente, nos gastos da secretaria da saúde.

O planejado para a gestão do SUS foi mantido, porém os gastos da atenção básica e os associados aos atendimentos de média e alta complexidade foram reduzidos pela metade. Serviços como visitas domiciliares e assistência à saúde bucal foram suspensos.

Outro fato relevante no período foi a decisão do governador de não cobrar tributos estaduais de algumas empresas para favorecer a implantação do novo parque industrial do leste do estado.

Assim como em outros estados, permanecem em debate os projetos de lei na assembleia legislativa para lidar com os prejuízos futuros da previdência do estado.

Figura 6 – Teste 2 B

No ano de 2017, o governo do estado teve uma frustração de receita de 27%. O planejamento de várias secretarias foi afetado, porém esse fato se refletiu, principalmente, nos gastos da secretaria da saúde.

O planejado para a gestão do SUS foi mantido, porém os gastos da atenção básica e os associados aos atendimentos de média e alta complexidade foram reduzidos pela metade. Serviços como visitas domiciliares e assistência à saúde bucal foram suspensos.

Outro fato relevante no período foi a renúncia de receita aprovada pelo prefeito no início do ano, favorecendo a implantação do novo parque industrial do leste do estado.

Assim como em outros estados, permanecem em debate os projetos de lei na assembleia legislativa para lidar com o aumento do passivo atuarial da previdência do estado.

Figura 7 – Teste 3 A

Balço do Patrimônio Público Geral do Município

fim do ano de 2017 (Em milhares de R\$)

Disponibilidades, bens e direitos			Obrigações		
	Ano atual	Ano anterior		Ano atual	Ano anterior
De curto prazo	R\$ 288,000.00	R\$ 174,000.00	De curto prazo	R\$ 211,000.00	R\$ 183,000.00
Investimentos financeiros disponíveis	R\$ 199,000.00	R\$ 135,000.00	Salários a pagar	R\$ 128,000.00	R\$ 102,000.00
Tributos a receber	R\$ 77,000.00	R\$ 34,000.00	Depósitos judiciais e consignações	R\$ 83,000.00	R\$ 81,000.00
Tributos atrasados a receber	R\$ 12,000.00	R\$ 5,000.00	De longo prazo	R\$ 266,000.00	R\$ 194,000.00
De longo prazo	R\$ 52,000.00	R\$ 52,000.00	Gastos futuros com aposentadorias, pensões e outros	R\$ 108,000.00	R\$ 85,000.00
Participação em outras empresas públicas	R\$ 46,000.00	R\$ 46,000.00	Empréstimos e financiamentos	R\$ 158,000.00	R\$ 109,000.00
Softwares	R\$ 6,000.00	R\$ 6,000.00	Saldo do Patrimônio Público	-R\$ 137,000.00	-R\$ 151,000.00
Total	R\$ 340,000.00	R\$ 226,000.00	Total	R\$ 340,000.00	R\$ 226,000.00

Qual a situação financeira da prefeitura?

Figura 8 – Teste 3 B

Balço Patrimonial Geral do Município

fechamento do exercício 2017 (Em milhares de R\$)

ATIVO			PASSIVO		
	Exercício atual	Exercício Anterior		Exercício atual	Exercício Anterior
Ativo circulante	R\$ 288,000.00	R\$ 174,000.00	Passivo circulante	R\$ 211,000.00	R\$ 183,000.00
Caixa e equivalente de caixa	R\$ 199,000.00	R\$ 135,000.00	Pessoal a pagar	R\$ 128,000.00	R\$ 102,000.00
Créditos tributários a receber	R\$ 77,000.00	R\$ 34,000.00	Valores restituíveis	R\$ 83,000.00	R\$ 81,000.00
Dívida ativa tributária	R\$ 12,000.00	R\$ 5,000.00	Passivo não-circulante	R\$ 266,000.00	R\$ 194,000.00
Ativo não-circulante	R\$ 52,000.00	R\$ 52,000.00	Provisão Matemática Previdenciária	R\$ 108,000.00	R\$ 85,000.00
Participações permanentes	R\$ 46,000.00	R\$ 46,000.00	Empréstimos e financiamentos	R\$ 158,000.00	R\$ 109,000.00
Intangível	R\$ 6,000.00	R\$ 6,000.00	Patrimônio Líquido	-R\$ 137,000.00	-R\$ 151,000.00
Total	R\$ 340,000.00	R\$ 226,000.00	Total	\$ 340,000.00	R\$ 226,000.00

Qual a situação financeira da prefeitura?

APÊNDICE C – Perguntas atenção e compreensão testes 1 e 2

<p>Teste 1 - Questão 1 (t1q1)</p> <div data-bbox="244 490 871 741"> <p>As obras das escolas contaram com dinheiro de:</p> <ul style="list-style-type: none"> ☐ Tributos. ☐ Doação. ☐ Empréstimo. ☐ Não Sei. <p>Next</p> </div>	<p>Resposta correta: Doação.</p> <p>Escala: 1 para correto, 0 para outras respostas.</p>
<p>Teste 1 - Questão 2 (t1q2)</p> <div data-bbox="244 880 813 1137"> <p>O texto trata de:</p> <ul style="list-style-type: none"> ☐ Um governo que faz um empréstimo para financiar obras públicas. ☐ Um governo que só opera com recursos próprios. ☐ Do aumento de tributos. ☐ De iniciativas da secretaria de educação. ☐ Não sei. <p>Next</p> </div>	<p>Resposta correta: “um governo que faz ...”.</p> <p>Escala: 1 para correto, 0 para outras respostas.</p>
<p>Teste 2 - Questão 1 (t2q1)</p> <div data-bbox="233 1296 818 1579"> <p>Os serviços suspensos foram:</p> <ul style="list-style-type: none"> ☐ Atendimentos de alta complexidade. ☐ Assistência à saúde bucal e visitas domiciliares. ☐ Gestão do SUS. ☐ Campanha de Vacinação. ☐ Não Sei. <p>Next</p> </div>	<p>Resposta correta: “Assistência a saúde bucal e visitas domiciliares”.</p> <p>Escala: 1 para correto, 0 para outras respostas.</p>
<p>Teste 2 - Questão 2 (t2q2)</p> <div data-bbox="233 1684 815 1966"> <p>O texto trata de:</p> <ul style="list-style-type: none"> ☐ Um governo que teve problema com a arrecadação de receitas. ☐ Um município que está aumentando impostos para melhorar a arrecadação. ☐ Dos projetos de lei para lidar com os problemas da previdência dos funcionários estado. ☐ Novas iniciativas da Secretaria de Saúde. ☐ Não sei. <p>Next</p> </div>	<p>Resposta correta: “Um governo que teve problema com a arrecadação de receitas”.</p> <p>Escala: 1 para correto, 0 para outras respostas.</p>

APÊNDICE D – Mapas de calor dos testes T1 a T3

Teste 1 A – maior legibilidade (n = 24)

a de infraestrutura faz a manutenção dos bens da prefeitura, além das obras públicas. Em 2017, o valor definido no orçamento para esta secretaria foi r, totalizando 33 milhões de reais. Parte das despesas foi destinada ao de melhoria da infraestrutura e mobilidade urbana. No programa, estão incluídos o recapeando de ruas e avenidas e reforma dos pontos de ônibus. Além de obras para melhorar a acessibilidade dos edifícios da prefeitura. De acordo com as demonstrações contábeis, o passivo aumentou 20% em relação a 2016. Esse aumento é consequência de um empréstimo obtido pela prefeitura. A verba foi usada para financiar as obras do programa de mobilidade urbana.

Além disso, a secretaria terminou de construir de três escolas cujas obras estavam paradas há cinco anos. Essa medida só foi possível usando receitas que não haviam sido previstas em orçamento. O dinheiro veio de doações recebidas de grandes empresários da região.

Teste 1 B – (n = 26)

secretaria de infraestrutura é responsável pela manutenção dos bens da prefeitura, além de fazer as obras e reformas públicas. No ano de 2017, o valor que foi separado no orçamento para esta secretaria foi no montante de 33 milhões de reais, 10% a mais que 2016, sendo que parte das despesas foi destinada ao programa de melhoria da infraestrutura e mobilidade urbana. Entre as medidas do programa está incluso o recapeamento de ruas e avenidas, reforma dos pontos de ônibus e obras para melhorar a acessibilidade dos edifícios da prefeitura. De acordo com as demonstrações contábeis, o passivo aumentou 20% em relação a 2016, esse aumento é consequência de um empréstimo obtido pela prefeitura para financiar as obras do programa de mobilidade urbana citadas.

Além disso, a secretaria finalizou a construção de três escolas que tinham obras paradas há cinco anos refletindo em um aumento no ativo do município, isso só foi possível pois o governo contou com receitas que não haviam sido previstas em orçamento, advindas de doações recebidas de grandes empresários da região.

Teste 2 A (n = 26)

No ano de 2017, a receita arrecadada foi 27% menor do que a receita que havia sido prevista. O planejamento de várias secretarias foi afetado, porém esse fato se refletiu, principalmente, nos gastos da secretaria da saúde.

O planejado para a gestão do SUS foi mantido, porém os gastos da atenção básica e os associados aos atendimentos de média e alta complexidade foram reduzidos pela metade. Serviços como visitas domiciliares e assistência à saúde bucal foram suspensos.

Outro fato relevante no período foi a decisão do governador de não cobrar tributos estaduais de algumas empresas para favorecer a implantação do novo parque industrial do leste do estado.

Assim como em outros estados, permanecem em debate os projetos de lei na assembleia legislativa para lidar com os prejuízos futuros da previdência do estado.

Teste 2 B (n = 24)

No ano de 2017, o governo do estado teve uma frustração de receita de 27%. O planejamento de várias secretarias foi afetado, porém esse fato se refletiu, principalmente, nos gastos da secretaria da saúde.

O planejado para a gestão do SUS foi mantido, porém os gastos da atenção básica e os associados aos atendimentos de média e alta complexidade foram reduzidos pela metade. Serviços como visitas domiciliares e assistência à saúde bucal foram suspensos.

Outro fato relevante no período foi a renúncia de receita aprovada pelo prefeito no início do ano, favorecendo a implantação do novo parque industrial do leste do estado.

Assim como em outros estados, permanecem em debate os projetos de lei na assembleia legislativa para lidar com o aumento do passivo atuarial da previdência do estado.

Teste 3A (n = 24)

Balanco do Patrimônio Público Geral do Município

fim do ano de 2017 (Em milhares de R\$)

Disponibilidades, bens e direitos			Obrigações		
	Ano atual	Ano anterior		Ano atual	Ano anterior
De curto prazo	R\$ 288,000.00	R\$ 174,000.00	De curto prazo	R\$ 211,000.00	R\$ 183,000.00
Investimentos financeiros disponíveis	R\$ 199,000.00	R\$ 135,000.00	Salários a pagar	R\$ 128,000.00	R\$ 102,000.00
Tributos a receber	R\$ 77,000.00	R\$ 34,000.00	Depósitos judiciais e consignações	R\$ 83,000.00	R\$ 81,000.00
Tributos atrasados a receber	R\$ 12,000.00	R\$ 5,000.00	De longo prazo	R\$ 266,000.00	R\$ 194,000.00
De longo prazo	R\$ 52,000.00	R\$ 52,000.00	Gastos futuros com aposentadorias, pensões e outros	R\$ 108,000.00	R\$ 85,000.00
Participação em outras empresas públicas	R\$ 46,000.00	R\$ 46,000.00	Empréstimos e financiamentos	R\$ 158,000.00	R\$ 109,000.00
Softwares	R\$ 6,000.00	R\$ 6,000.00	Saldo do Patrimônio Público	-R\$ 137,000.00	-R\$ 151,000.00
Total	R\$ 340,000.00	R\$ 226,000.00	Total	R\$ 340,000.00	R\$ 226,000.00

Qual a situação financeira da prefeitura?

Teste 3 B (n = 26)

Balanco Patrimonial Geral do Município

fechamento do exercício 2017 (Em milhares de R\$)

ATIVO			PASSIVO		
	Exercício atual	Exercício Anterior		Exercício atual	Exercício Anterior
Ativo circulante	R\$ 288,000.00	R\$ 174,000.00	Passivo circulante	R\$ 211,000.00	R\$ 183,000.00
Caixa e equivalente de caixa	R\$ 199,000.00	R\$ 135,000.00	Pessoal a pagar	R\$ 128,000.00	R\$ 102,000.00
Créditos tributários a receber	R\$ 77,000.00	R\$ 34,000.00	Valores restituíveis	R\$ 83,000.00	R\$ 81,000.00
Dívida ativa tributária	R\$ 12,000.00	R\$ 5,000.00	Passivo não-circulante	R\$ 266,000.00	R\$ 194,000.00
Ativo não-circulante	R\$ 52,000.00	R\$ 52,000.00	Provisão Matemática Previdenciária	R\$ 108,000.00	R\$ 85,000.00
Participações permanentes	R\$ 46,000.00	R\$ 46,000.00	Empréstimos e financiamentos	R\$ 158,000.00	R\$ 109,000.00
Intangível	R\$ 6,000.00	R\$ 6,000.00	Patrimônio Líquido	-R\$ 137,000.00	-R\$ 151,000.00
Total	R\$ 340,000.00	R\$ 226,000.00	Total	\$ 340,000.00	R\$ 226,000.00

Qual a situação financeira da prefeitura?

APÊNDICE E – Tabelas e mais dados com análises descritivas e estatísticas

Todas as análises estatísticas, gráficos e tabelas foram geradas usando o software StataSE 12. Todas as figuras aqui apresentadas são de elaboração do autor.

1. (T1) Teste de Shapiro-Wilk para averiguar se a distribuição dos dados é normal. A probabilidade Prob>z da última coluna deve ser maior que 0.05.

```
. swilk t1q1 t1q2 t1q3 t1q4 t2q1 t2q2 t2q3 t2q4 crt1 crt2 crt3 crt4
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
t1q1	50	0.95854	1.950	1.424	0.07722
t1q2	50	0.99331	0.315	-2.465	0.99315
t1q3	50	0.97380	1.232	0.446	0.32795
t1q4	50	0.98301	0.799	-0.479	0.68403
t2q1	50	0.70489	13.879	5.610	0.00000
t2q2	50	0.70489	13.879	5.610	0.00000
t2q3	50	0.95802	1.974	1.451	0.07342
t2q4	50	0.97215	1.310	0.575	0.28260
crt1	50	0.98793	0.568	-1.207	0.88627
crt2	50	0.97295	1.272	0.513	0.30400
crt3	50	0.60802	18.435	6.215	0.00000
crt4	50	0.96280	1.749	1.193	0.11647

2. (T1) Estatísticas de média simples para as questões 1 e 2 (t1q1 e t1q2).

Summary statistics: N, mean
by categories of: tipo (Se fez o caminho A ou B)

tipo	t1q1	t1q2
A	24 .875	24 .5
B	26 .5769231	26 .6923077
Total	50 .72	50 .6

3. (T1) Teste T para t1q1 por tipo (se seguiu caminho 1 ou 2)

```
. ttest t1q1, by (tipo)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
A	24	.875	.0689597	.337832	.7323461	1.017654
B	26	.5769231	.0988095	.5038315	.3734211	.780425
combined	50	.72	.0641427	.4535574	.5911004	.8488996
diff		.2980769	.122376		.0520235	.5441304

diff = mean(A) - mean(B) t = 2.4357
 Ho: diff = 0 degrees of freedom = 48

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.9907 Pr(|T| > |t|) = 0.0186 Pr(T > t) = 0.0093

4. (T1) Teste T para t1q2 por tipo (se seguiu caminho 1 ou 2)

```
. ttest t1q2, by (tipo)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
A	24	.5	.1042572	.5107539	.2843275	.7156725
B	26	.6923077	.0923077	.4706787	.5021964	.8824189
combined	50	.6	.0699854	.4948717	.459359	.740641
diff		-.1923077	.1387861		-.4713559	.0867405

diff = mean(A) - mean(B) t = -1.3856
 Ho: diff = 0 degrees of freedom = 48

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.0861 Pr(|T| > |t|) = 0.1723 Pr(T > t) = 0.9139

5. (T1) Médias simples para avaliar se respondentes que acertam pergunta 1 apresentam melhor desempenho na pergunta 2. Primeira tabela t1q1==0, segunda t1q1==1.

Summary statistics: N, mean by categories of: tipo (Se fez o caminho A)			Summary statistics: N, mean by categories of: tipo (Se fez o caminho A)		
tipo	t1q1	t1q2	tipo	t1q1	t1q2
A	3	3	A	21	21
	0	.6666667		1	.4761905
B	11	11	B	15	15
	0	.6363636		1	.7333333
Total	14	14	Total	36	36
	0	.6428571		1	.5833333

6. (T1) Teste T para averiguar se respondentes que acertam pergunta 1 estão mais sujeitos aos efeitos da manipulação. Resultado deu não significativo, ao contrário do que se esperava.

```
. ttest    t1q2 if t1q1==1, by(tipo)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
A	21	.4761905	.1116766	.5117663	.2432372	.7091437
B	15	.7333333	.1181874	.4577377	.4798466	.98682
combined	36	.5833333	.0833333	.5	.4141577	.752509
diff		-.2571429	.1657317		-.5939501	.0796644

[illegible]

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.0650	Pr(T > t) = 0.1300	Pr(T > t) = 0.9350

7. (T1) Teste de Shapiro-Wilk para averiguar se a distribuição dos dados das métricas geradas pelo *eye-tracker* para AOI a7 do teste 1A e 1B é normal. A probabilidade Prob>z da última coluna deve ser maior que 0.05. A única variável com distribuição normal é Fixação Total (pois Prob>z de ft_t1ab_a7 > 0.05).

```
. swilk  fm tlab a7 fc tlab a7 ft tlab a7 vm tlab a7 vc tlab a7 vt tlab a7 dur tlab
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
fm_tlab_a7	50	0.90048	4.680	3.291	0.00050
fc_tlab_a7	50	0.92069	3.730	2.807	0.00250
ft_tlab_a7	50	0.96241	1.768	1.215	0.11212
vm_tlab_a7	50	0.82350	8.301	4.513	0.00000
vc_tlab_a7	50	0.78569	10.079	4.927	0.00000
vt_tlab_a7	50	0.91456	4.018	2.966	0.00151
dur_tlab	50	0.92657	3.453	2.643	0.00411

8. (T1) Teste T com a variável `ft_tlab_a7` para averiguar se há diferença significativas entre as médias das amostras que tiveram acesso ao texto facilitado e as que não tiveram.

```
. ttest ft_tlab_a7, by (tipo)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
A	24	71.35917	4.821513	23.6205	61.38511	81.33323
B	26	67.86192	4.640686	23.66295	58.30425	77.4196
combined	50	69.5406	3.318677	23.46659	62.87147	76.20973
diff		3.497243	6.692498		-9.958926	16.95341

```
diff = mean(A) - mean(B)                                t = 0.5226
Ho: diff = 0                                             degrees of freedom = 48
```

```
Ha: diff < 0                Ha: diff != 0                Ha: diff > 0
Pr(T < t) = 0.6982          Pr(|T| > |t|) = 0.6037          Pr(T > t) = 0.3018
```

9. (T1) Testes de Mann-Whitney (não paramétrico) para as métricas geradas pelo *Eye-tracker* sem distribuição normal – `fm_tlab_a7`. Por último, tabela com média, min, max e desvio padrão.

```
. ranksum fm_tlab_a7, by(tipo) porder

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      tipo |      obs   rank sum   expected
-----|-----
      A    |      24       681       612
      B    |      26       594       663
-----|-----
 combined  |      50      1275      1275

unadjusted variance      2652.00
adjustment for ties      -18.59
-----
adjusted variance        2633.41

Ho: fm~ab_a7(tipo==A) = fm~ab_a7(tipo==B)
      z = 1.345
Prob > |z| = 0.1788

P{fm~ab_a7(tipo==A) > fm~ab_a7(tipo==B)} = 0.611
```

```
. ranksum fc_tlab_a7, by(tipo) porder

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      tipo |      obs   rank sum   expected
-----|-----
      A    |      24      636.5       612
      B    |      26      638.5       663
-----|-----
 combined  |      50      1275      1275

unadjusted variance      2652.00
adjustment for ties      -1.78
-----
adjusted variance        2650.22

Ho: fc~ab_a7(tipo==A) = fc~ab_a7(tipo==B)
      z = 0.476
Prob > |z| = 0.6341

P{fc~ab_a7(tipo==A) > fc~ab_a7(tipo==B)} = 0.539
```

```
. ranksum vt_tlab_a7, by(tipo) porder

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      tipo |      obs   rank sum   expected
-----|-----
      A    |      24       628       612
      B    |      26       647       663
-----|-----
 combined  |      50      1275      1275

unadjusted variance      2652.00
adjustment for ties        0.00
-----
adjusted variance        2652.00

Ho: vt~ab_a7(tipo==A) = vt~ab_a7(tipo==B)
      z = 0.311
Prob > |z| = 0.7560

P{vt~ab_a7(tipo==A) > vt~ab_a7(tipo==B)} = 0.526
```

```
. ranksum vm_tlab_a7, by(tipo) porder

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      tipo |      obs   rank sum   expected
-----|-----
      A    |      24       585       612
      B    |      26       690       663
-----|-----
 combined  |      50      1275      1275

unadjusted variance      2652.00
adjustment for ties        0.00
-----
adjusted variance        2652.00

Ho: vm~ab_a7(tipo==A) = vm~ab_a7(tipo==B)
      z = -0.524
Prob > |z| = 0.6001

P{vm~ab_a7(tipo==A) > vm~ab_a7(tipo==B)} = 0.457
```

. ranksum vc_tlab_a7, by(tipo) porder				. ranksum dur_tlab, by(tipo) porder			
Two-sample Wilcoxon rank-sum (Mann-Whitney) test				Two-sample Wilcoxon rank-sum (Mann-Whitney) test			
tipo	obs	rank sum	expected	tipo	obs	rank sum	expected
A	24	645.5	612	A	24	634	612
B	26	629.5	663	B	26	641	663
combined	50	1275	1275	combined	50	1275	1275
unadjusted variance		2652.00		unadjusted variance		2652.00	
adjustment for ties		-55.40		adjustment for ties		0.00	
adjusted variance		2596.60		adjusted variance		2652.00	
Ho: vc~ab_a7(tipo==A) = vc~ab_a7(tipo==B)				Ho: dur_tlab(tipo==A) = dur_tlab(tipo==B)			
z = 0.657				z = 0.427			
Prob > z = 0.5109				Prob > z = 0.6692			
P{vc~ab_a7(tipo==A) > vc~ab_a7(tipo==B)} = 0.55				P{dur_tlab(tipo==A) > dur_tlab(tipo==B)} = 0.53			
Summary statistics: mean, max, min, sd by categories of: tipo (Se fez o caminho A ou B)							
tipo	fm~ab_a7	fc~ab_a7	ft~ab_a7	vm~ab_a7	vc~ab_a7	vt~ab_a7	dur_tlab
A	.3866667	183.5	71.35917	30.52292	5.9375	79.72583	84.8225
	.46	338	116.38	79.71	21	145.72	152.31
	.16	51	8.39	3.75	1	40.69	45.68
	.0614793	64.93944	23.6205	24.83256	5.838985	25.00822	28.15662
B	.3730769	181.6154	67.86192	33.19346	4.096154	77.79269	82.57461
	.51	340	130.78	147.53	16	147.53	147.53
	.21	109	24.31	5.87	1	41.25	41.62
	.0550468	58.4986	23.66295	30.07821	3.644226	24.13065	25.49592
Total	.3796	182.52	69.5406	31.9116	4.98	78.7206	83.6536
	.51	340	130.78	147.53	21	147.53	152.31
	.16	51	8.39	3.75	1	40.69	41.62
	.0580274	61.04372	23.46659	27.43808	4.86235	24.32279	26.55315

10. (T2) Médias simples das respostas às perguntas de compreensão do Teste 2 e teste de normalidade da distribuição dos dados.

F(111);				. swilk t2q1 t2q2 t2q3 t2q4 t2q1r t2q2r			
. tabstat t2q1 t2q2, by (tipo) stats (mean, max, min, sd)				Shapiro-Wilk W test for normal data			
Summary statistics: mean, max, min, sd by categories of: tipo (Se fez o caminho A ou B)				Variable	Obs	W	V
tipo	t2q1	t2q2			z	Prob	
A	.8333333	.9166667		t2q1	50	0.70489	13.879
	1	1		t2q2	50	0.70489	13.879
	0	0		t2q3	50	0.95802	1.974
	.3806935	.2823299		t2q4	50	0.97215	1.310
B	.9615385	.8846154		t2q1r	50	0.83167	7.916
	1	1		t2q2r	50	0.74583	11.954
	0	0					
	.1961161	.3258126					
Total	.9	.9					
	1	1					
	0	0					
	.3030458	.3030458					

11. (T2) Testes de Mann-Whitney para análises das perguntas de compreensão do Teste 2.

<pre>. ranksum t2q1, by(tipo) porder</pre> <p>Two-sample Wilcoxon rank-sum (Mann-Whitney) test</p> <table><thead><tr><th>tipo</th><th>obs</th><th>rank sum</th><th>expected</th></tr></thead><tbody><tr><td>A</td><td>24</td><td>572</td><td>612</td></tr><tr><td>B</td><td>26</td><td>703</td><td>663</td></tr><tr><td>combined</td><td>50</td><td>1275</td><td>1275</td></tr></tbody></table> <p>unadjusted variance 2652.00 adjustment for ties -1935.67</p> <hr/> <p>adjusted variance 716.33</p> <p>Ho: t2q1(tipo==A) = t2q1(tipo==B) z = -1.495 Prob > z = 0.1350</p> <p>P{t2q1(tipo==A) > t2q1(tipo==B)} = 0.436</p>	tipo	obs	rank sum	expected	A	24	572	612	B	26	703	663	combined	50	1275	1275	<pre>. ranksum t2q2, by(tipo) porder</pre> <p>Two-sample Wilcoxon rank-sum (Mann-Whitney) test</p> <table><thead><tr><th>tipo</th><th>obs</th><th>rank sum</th><th>expected</th></tr></thead><tbody><tr><td>A</td><td>24</td><td>622</td><td>612</td></tr><tr><td>B</td><td>26</td><td>653</td><td>663</td></tr><tr><td>combined</td><td>50</td><td>1275</td><td>1275</td></tr></tbody></table> <p>unadjusted variance 2652.00 adjustment for ties -1935.67</p> <hr/> <p>adjusted variance 716.33</p> <p>Ho: t2q2(tipo==A) = t2q2(tipo==B) z = 0.374 Prob > z = 0.7087</p> <p>P{t2q2(tipo==A) > t2q2(tipo==B)} = 0.516</p>	tipo	obs	rank sum	expected	A	24	622	612	B	26	653	663	combined	50	1275	1275
tipo	obs	rank sum	expected																														
A	24	572	612																														
B	26	703	663																														
combined	50	1275	1275																														
tipo	obs	rank sum	expected																														
A	24	622	612																														
B	26	653	663																														
combined	50	1275	1275																														
<pre>. ranksum t2q2 if t1q1==1, by(tipo) porder</pre> <p>Two-sample Wilcoxon rank-sum (Mann-Whitney) test</p> <table><thead><tr><th>tipo</th><th>obs</th><th>rank sum</th><th>expected</th></tr></thead><tbody><tr><td>A</td><td>21</td><td>412.5</td><td>388.5</td></tr><tr><td>B</td><td>15</td><td>253.5</td><td>277.5</td></tr><tr><td>combined</td><td>36</td><td>666</td><td>666</td></tr></tbody></table> <p>unadjusted variance 971.25 adjustment for ties -683.25</p> <hr/> <p>adjusted variance 288.00</p> <p>Ho: t2q2(tipo==A) = t2q2(tipo==B) z = 1.414 Prob > z = 0.1573</p> <p>P{t2q2(tipo==A) > t2q2(tipo==B)} = 0.576</p>	tipo	obs	rank sum	expected	A	21	412.5	388.5	B	15	253.5	277.5	combined	36	666	666																	
tipo	obs	rank sum	expected																														
A	21	412.5	388.5																														
B	15	253.5	277.5																														
combined	36	666	666																														

12. (T2) Teste de normalidade (Shapiro Wilk) métricas do teste 2. As únicas que tem distribuição normal são fc e ft. As outras métricas não.


```
. ranksum  vc_t2ab_a5, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	24	723	612
B	26	552	663
combined	50	1275	1275

unadjusted variance 2652.00
adjustment for ties -37.82
adjusted variance 2614.18

Ho: vc~ab_a5(tipo==A) = vc~ab_a5(tipo==B)
z = 2.171
Prob > |z| = 0.0299

P{vc~ab_a5(tipo==A) > vc~ab_a5(tipo==B)} = 0.678

```
. ranksum  vt_t2ab_a5, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	24	693	612
B	26	582	663
combined	50	1275	1275

unadjusted variance 2652.00
adjustment for ties 0.00
adjusted variance 2652.00

Ho: vt~ab_a5(tipo==A) = vt~ab_a5(tipo==B)
z = 1.573
Prob > |z| = 0.1157

P{vt~ab_a5(tipo==A) > vt~ab_a5(tipo==B)} = 0.630

```
. ranksum  dur_t2ab, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	24	705	612
B	26	570	663
combined	50	1275	1275

unadjusted variance 2652.00
adjustment for ties 0.00
adjusted variance 2652.00

Ho: dur_t2ab(tipo==A) = dur_t2ab(tipo==B)
z = 1.806
Prob > |z| = 0.0709

P{dur_t2ab(tipo==A) > dur_t2ab(tipo==B)} = 0.649

14. (T2) Teste Shapiro-Wilk para averiguar normalidade na distribuição da amostra

<pre>. swilk fm_t2ab_a1 fc_t2ab_a1 ft_t2ab_a1 vm_t2ab_a1 vc_t2ab_a1 vt_t2ab_a1</pre>					
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
fm_t2ab_a1	41	0.75328	9.940	4.840	0.00000
fc_t2ab_a1	41	0.83557	6.624	3.985	0.00003
ft_t2ab_a1	41	0.85071	6.014	3.781	0.00008
vm_t2ab_a1	41	0.88372	4.684	3.255	0.00057
vc_t2ab_a1	41	0.82184	7.177	4.154	0.00002
vt_t2ab_a1	41	0.84719	6.156	3.830	0.00006

16. (T2) Teste Shapiro-Wilk para averiguar normalidade na distribuição da amostra

```
. label variable vt_t2ab_a2 "junção de vt a e b para a2"

. swilk fm_t2ab_a2 fc_t2ab_a2 ft_t2ab_a2 vm_t2ab_a2 vc_t2ab_a2 vt_t2ab_a2
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
fm_t2ab_a2	37	0.97502	0.930	-0.152	0.56024
fc_t2ab_a2	37	0.89677	3.844	2.820	0.00240
ft_t2ab_a2	37	0.85978	5.221	3.462	0.00027
vm_t2ab_a2	37	0.90945	3.372	2.546	0.00545
vc_t2ab_a2	37	0.91664	3.104	2.372	0.00884
vt_t2ab_a2	37	0.86219	5.131	3.425	0.00031

17. (T2) Ttest para fm e Mann-Whitney para as variáveis fc, ft, vm, vc, e vt de a2 (jargão 2)

```
. ttest fm_t2ab_a2, by (tipo)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
A	18	.3966667	.0338006	.143404	.3253535 .4679798
B	19	.4015789	.0356006	.1551796	.3267848 .4763731
combined	37	.3991892	.0242496	.1475048	.3500086 .4483697
diff		-.0049123	.0491982		-.1047899 .0949654

diff = mean(A) - mean(B) t = -0.0998
Ho: diff = 0 degrees of freedom = 35

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.4605 Pr(|T| > |t|) = 0.9210 Pr(T > t) = 0.5395

```
. ranksum ft_t2ab_a2, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	18	401	342
B	19	302	361
combined	37	703	703

unadjusted variance 1083.00
adjustment for ties -0.26

adjusted variance 1082.74

Ho: ft~ab_a2(tipo==A) = ft~ab_a2(tipo==B)

z = 1.793

Prob > |z| = 0.0730

P{ft~ab_a2(tipo==A) > ft~ab_a2(tipo==B)} = 0.673

```
. ranksum fc_t2ab_a2, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	18	414.5	342
B	19	288.5	361
combined	37	703	703

unadjusted variance 1083.00
adjustment for ties -28.37

adjusted variance 1054.63

Ho: fc~ab_a2(tipo==A) = fc~ab_a2(tipo==B)

z = 2.232

Prob > |z| = 0.0256

P{fc~ab_a2(tipo==A) > fc~ab_a2(tipo==B)} = 0.712

```
. ranksum vt_t2ab_a2, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	18	394	342
B	19	309	361
combined	37	703	703

unadjusted variance 1083.00
adjustment for ties -0.13

adjusted variance 1082.87

Ho: vt~ab_a2(tipo==A) = vt~ab_a2(tipo==B)

z = 1.580

Prob > |z| = 0.1141

P{vt~ab_a2(tipo==A) > vt~ab_a2(tipo==B)} = 0.652

```
. ranksum  vm_t2ab_a2, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	18	358	342
B	19	345	361
combined	37	703	703

unadjusted variance 1083.00
adjustment for ties -0.26

adjusted variance 1082.74

Ho: vm~ab_a2(tipo==A) = vm~ab_a2(tipo==B)
z = 0.486
Prob > |z| = 0.6268

P{vm~ab_a2(tipo==A) > vm~ab_a2(tipo==B)} = 0.547

```
. ranksum  vc_t2ab_a2, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	18	414	342
B	19	289	361
combined	37	703	703

unadjusted variance 1083.00
adjustment for ties -91.02

adjusted variance 991.98

Ho: vc~ab_a2(tipo==A) = vc~ab_a2(tipo==B)
z = 2.286
Prob > |z| = 0.0223

P{vc~ab_a2(tipo==A) > vc~ab_a2(tipo==B)} = 0.711

18. (T2) Teste Shapiro-Wilk para averiguar normalidade na distribuição da amostra de a3

```
. swilk fm_t2ab_a3 fc_t2ab_a3 ft_t2ab_a3 vm_t2ab_a3 vc_t2ab_a3 vt_t2ab_a3
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
fm_t2ab_a3	42	0.96733	1.341	0.619	0.26788
fc_t2ab_a3	42	0.80176	8.137	4.425	0.00000
ft_t2ab_a3	42	0.79053	8.598	4.541	0.00000
vm_t2ab_a3	42	0.94488	2.262	1.723	0.04244
vc_t2ab_a3	42	0.80919	7.832	4.344	0.00001
vt_t2ab_a3	42	0.78183	8.955	4.627	0.00000

19. (T2) Ttest para fm e Mann-Whitney para as variáveis fc, ft, vm, vc, e vt de a2 (jargão 2)

<pre>. ttest fm_t2ab_a3, by (tipo)</pre>						<pre>. ranksum fc_t2ab_a3, by (tipo) porder</pre>			
Two-sample t test with equal variances						Two-sample Wilcoxon rank-sum (Mann-Whitney) test			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	tipo	obs	rank sum	expected
A	21	.3909524	.0340511	.1560418	.319923 .4619818	A	21	488	451.5
B	21	.3571429	.0256082	.1173517	.303725 .4105607	B	21	415	451.5
combined	42	.3740476	.0212065	.137434	.3312202 .4168751	combined	42	903	903
diff		.0338095	.0426059		-.0523002 .1199192	unadjusted variance 1580.25			
diff = mean(A) - mean(B)					t = 0.7935	adjustment for ties -9.73			
Ho: diff = 0					degrees of freedom = 40	adjusted variance 1570.52			
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		Ho: fc~ab_a3(tipo==A) = fc~ab_a3(tipo==B)			
Pr(T < t) = 0.7839		Pr(T > t) = 0.4321		Pr(T > t) = 0.2161		z = 0.921			
						Prob > z = 0.3570			
						P(fc~ab_a3(tipo==A) > fc~ab_a3(tipo==B)) = 0.583			

```
. ranksum ft_t2ab_a3, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	21	491	451.5
B	21	412	451.5
combined	42	903	903

unadjusted variance 1580.25
adjustment for ties -0.13
adjusted variance 1580.12

Ho: ft~ab_a3(tipo==A) = ft~ab_a3(tipo==B)
z = 0.994
Prob > |z| = 0.3204

P{ft~ab_a3(tipo==A) > ft~ab_a3(tipo==B)} = 0.590

```
. ranksum vc_t2ab_a3, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	21	498	451.5
B	21	405	451.5
combined	42	903	903

unadjusted variance 1580.25
adjustment for ties -28.55
adjusted variance 1551.70

Ho: vc~ab_a3(tipo==A) = vc~ab_a3(tipo==B)
z = 1.180
Prob > |z| = 0.2378

P{vc~ab_a3(tipo==A) > vc~ab_a3(tipo==B)} = 0.605

```
. ranksum vm_t2ab_a3, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	21	454	451.5
B	21	449	451.5
combined	42	903	903

unadjusted variance 1580.25
adjustment for ties -1.02
adjusted variance 1579.23

Ho: vm~ab_a3(tipo==A) = vm~ab_a3(tipo==B)
z = 0.063
Prob > |z| = 0.9498

P{vm~ab_a3(tipo==A) > vm~ab_a3(tipo==B)} = 0.506

```
. ranksum vt_t2ab_a3, by (tipo) porder
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	21	487	451.5
B	21	416	451.5
combined	42	903	903

unadjusted variance 1580.25
adjustment for ties -0.26
adjusted variance 1579.99

Ho: vt~ab_a3(tipo==A) = vt~ab_a3(tipo==B)
z = 0.893
Prob > |z| = 0.3718

P{vt~ab_a3(tipo==A) > vt~ab_a3(tipo==B)} = 0.580

20. (T3) Teste Shapiro-Wilk para averiguar normalidade na distribuição da amostra da AOI a3

```
. swilk fm_t3ab_a3 fc_t3ab_a3 ft_t3ab_a3 vm_t3ab_a3 vc_t3ab_a3 vt_t3ab_a3 fm_t3
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
fm_t3ab_a3	50	0.82875	8.054	4.449	0.00000
fc_t3ab_a3	50	0.94650	2.516	1.968	0.02455
ft_t3ab_a3	50	0.95515	2.109	1.592	0.05571
vm_t3ab_a3	50	0.78152	10.275	4.968	0.00000
vc_t3ab_a3	50	0.95965	1.898	1.366	0.08594
vt_t3ab_a3	50	0.95179	2.267	1.746	0.04045
fm_t3ab_a4	50	0.85663	6.743	4.070	0.00002
fc_t3ab_a4	50	0.98012	0.935	-0.144	0.55711
ft_t3ab_a4	50	0.98825	0.553	-1.265	0.89703
vm_t3ab_a4	50	0.91104	4.184	3.052	0.00114
vc_t3ab_a4	50	0.95432	2.148	1.631	0.05148
vt_t3ab_a4	50	0.97160	1.336	0.617	0.26853

21. (T3) Testes T para ver a diferença de médias entre as versões A e B (variáveis com distribuição normal)

ttest ft_t3ab_a3 , by (tipo) ; ttest vc_t3ab_a3 , by (tipo); ttest fc_t3ab_a4 , by (tipo) ;
ttest ft_t3ab_a4 , by (tipo); ttest vt_t3ab_a4 , by (tipo) => todos sem significância

22. (T3) Mann-Whitney para ver a diferença de médias entre as versões A e B (variáveis sem distribuição normal)

Ranksum (fm_t3ab_a3 /fc_t3ab_a3/ vm_t3ab_a3/ vt_t3ab_a3/ fm_t3ab_a4 /vm_t3ab_a4) , by (tipo) => todos sem significância.

```
. ranksum vm_t3ab_num_tex, by (tipo)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	24	666	612
B	26	609	663
combined	50	1275	1275

unadjusted variance 2652.00
adjustment for ties 0.00
adjusted variance 2652.00

Ho: vm_t3a..(tipo==A) = vm_t3a..(tipo==B)
z = 1.049
Prob > |z| = 0.2944

```
. ranksum ft_t3ab_num_tex, by (tipo)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	24	710	612
B	26	565	663
combined	50	1275	1275

unadjusted variance 2652.00
adjustment for ties 0.00
adjusted variance 2652.00

Ho: ft_t3a..(tipo==A) = ft_t3a..(tipo==B)
z = 1.903
Prob > |z| = 0.0570

```
. ranksum fc_t3ab_num_tex, by (tipo)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

tipo	obs	rank sum	expected
A	24	765.5	612
B	26	509.5	663
combined	50	1275	1275

unadjusted variance 2652.00
adjustment for ties -0.13
adjusted variance 2651.87

Ho: fc_t3a..(tipo==A) = fc_t3a..(tipo==B)
z = 2.981
Prob > |z| = 0.0029

23. (T3) Teste de normalidade variáveis criadas para análise de T3

```
. swilk ft_t3ab_num_tex ft_t3ab_atv_pas ft_t3ab_atu_ant vm_t3ab_num_
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
ft_t3ab_nu~x	50	0.91718	3.895	2.900	0.00187
ft_t3ab_at~s	50	0.78843	9.950	4.900	0.00000
ft_t3ab_at~t	49	0.79525	9.477	4.790	0.00000
vm_t3ab_nu~x	50	0.97995	0.943	-0.125	0.54974
vm_t3ab_at~s	50	0.86172	6.503	3.993	0.00003
vm_t3ab_at~t	49	0.85803	6.572	4.010	0.00003
fc_t3ab_nu~x	50	0.89228	5.066	3.460	0.00027
fc_t3ab_at~s	50	0.66428	15.789	5.885	0.00000
fc_t3ab_at~t	49	0.83368	7.699	4.348	0.00001

APÊNDICE F – Dados sobre a validação da dificuldade dos jargões contábeis

Tabela 16 – Termos técnicos da contabilidade e finanças públicas (jargões) classificados pelo autor e corroborado por especialistas na área de contabilidade pública

Área	n	Termos técnicos da contabilidade e finanças públicas	Total = 152
Nível de dificuldade de dificuldade: fácil			n = 46
econômica	6	recursos obtidos; recursos aplicados; arrecadação de tributos; endividamento estatal; gastos com manutenção; previdência	
financeiro	1	transferências financeiras	
fiscal	4	operações de crédito; índices; limite prudencial; produção, renda e consumo; dívida pública	
geral	8	prestação de contas anual; balanço orçamentário; despesas; exercício anterior; exercício atual; participação percentual; obtenção de recursos; exceto	
orçamentário	15	orçamento público; receita prevista; receita arrecadada; despesa fixada; Execução do orçamento; previsão atualizada; investimentos; amortização da dívida; juros e encargos da dívida; pessoal e encargos sociais; despesa orçada; despesa realizada; demonstrativo; superávit / déficit orçamentário; resultado orçamentário	
patrimonial	12	ativo; passivo; disponibilidades em caixa e equivalente de caixa; aplicações financeiras; clientes; estoques; créditos a curto prazo; empréstimos e financiamentos de curto prazo; fornecedores e contas a pagar CP; bens; direitos; obrigações	
Nível de dificuldade de dificuldade: médio			n = 67
econômica	9	conta redutora; recursos obtidos da exploração do patrimônio estatal; recursos obtidos de alienação de bens; demais receitas; outros serviços de terceiros; material de consumo; serviços da dívida; transferências constitucionais; outras transferências	
financeiro	7	balanço financeiro; dispêndios extra orçamentários; ingresso / dispêndio ordinário; ingresso / dispêndio vinculado; recebimentos / pagamentos extra orçamentários; saldo do exercício anterior; saldo para o exercício seguinte	
fiscal	9	renúncia de receita; dívida consolidada e mobiliária; antecipação de receita; concessão de garantia; demais gastos; dívida interna; dívida externa; encargos	
geral	8	tributos; imposto; taxa; contribuição de melhoria; empréstimo compulsório; contribuições sociais; administração direta; administração indireta	
orçamentário	11	despesa executada; exercício financeiro; receita orçamentária; receita orçamentária realizada / realização de receita; despesa orçamentária; inversões financeiras; despesa intra-orçamentária; despesa empenhada; receita frustrada / frustração de receita; dotação atualizada; dotação orçamentária	
patrimonial	23	créditos tributários a receber; obrigações trabalhistas; obrigações previdenciárias ; obrigações assistenciais a pagar; dívidas de empréstimos e financiamentos; ativo circulante; ativo não circulante; VPD pagas antecipadamente; ativo imobilizado; investimentos; demais obrigações a curto prazo; consignações recebidas; depósitos recebidos; convênios recebidos; obrigações trabalhistas, prev. e assistência a pagar a curto prazo; obrigações fiscais a CP; obrigações de repartição a outros entes; passivo não circulante; demais obrigações a longo prazo; obrigações trabalhistas previdenciárias e assistências a pagar a longo prazo; empréstimos e financiamentos LP; fornecedores LP; benefícios previdenciários	
Nível de dificuldade de dificuldade: difícil			n = 39
econômica	4	balanço econômico; superávit / déficit do balanço econômico; receita intraorçamentária	
financeiro	5	demonstração dos Fluxos de Caixa; fluxo de caixa das atividades operacionais; fluxo / fluxo líquido das atividades de investimentos	
fiscal	7	fluxo / fluxo líquido das atividades de financiamentos; geração líquida de caixa e equivalentes de caixa	
orçamentário	7	equilíbrio financeiro; Lei de Responsabilidade Fiscal LRF; inscrição em restos a pagar; dívida consolidada líquida; receita corrente líquida; resultado primário; resultado nominal	
		receita (orçamentária) corrente; receita (orçamentária) de capital; receita extra orçamentária; despesa corrente; despesa de capital; conta retificadora da receita corrente; gestão orçamentária	

patrimonial	16	provisões a longo prazo; balanço patrimonial; patrimônio líquido; dívida ativa tributária; participações permanentes; bens corpóreos; ativo intangível; ativo realizável a longo prazo; variação patrimonial ; superávit patrimonial; déficit patrimonial; demonstração das variações patrimoniais; variação patrimonial aumentativa (VPA); Variações Patrimoniais Diminutivas (VPD); provisão matemática previdenciária / passivo atuarial do Estado; déficit previdenciário
-------------	----	---

Fonte: elaborado pelo autor

Tabela 17 – Lista dos termos validados entre a população amostral

Termo técnico	Class. prévia	Class. validação	n	Média *	D.p.	Máx.	Mín.
dotação atualizada	médio	difícil	51	1.14	0.35	2	1
dispêndio ordinário	médio	difícil	51	1.16	0.42	3	1
passivo atuarial	difícil	difícil	51	1.35	0.56	3	1
renúncia de receita	médio	difícil	47	1.49	0.69	4	1
frustração de receita	médio	difícil	51	1.61	0.83	4	1
variação patrimonial aumentativa	difícil	difícil	51	1.61	0.80	4	1
receita intraorçamentária	difícil	difícil	51	1.61	0.72	4	1
consignações recebidas	médio	difícil	47	1.62	0.77	4	1
despesa executada	médio	difícil	51	1.94	0.95	4	1
receita orçamentária de capital	difícil	médio	47	2.00	0.91	4	1
dívida ativa tributária	difícil	médio	47	2.06	0.87	4	1
despesa de capital	difícil	médio	47	2.15	0.91	4	1
despesa corrente	difícil	médio	47	2.23	1.00	4	1
receita corrente	difícil	médio	51	2.45	0.99	4	1
gestão orçamentária	difícil	médio	51	2.59	0.96	4	1
exercício financeiro	médio	médio	51	2.65	1.13	4	1
participação percentual	fácil	médio	47	2.79	1.14	4	1
receita prevista	fácil	médio	51	2.80	1.06	4	1
exceto	fácil	médio	47	2.85	1.40	4	1
equilíbrio financeiro	difícil	médio	47	3.00	0.91	4	1
passivo não circulante	médio	fácil	47	3.02	1.26	4	1
benefícios previdenciários	médio	fácil	47	3.04	0.88	4	1
ativo circulante	médio	fácil	47	3.15	1.20	4	1
déficit previdenciário	difícil	fácil	47	3.19	0.99	4	1
índices	fácil	fácil	47	3.23	0.89	4	1
prestação de contas anual	fácil	fácil	47	3.28	0.88	4	1
ativo realizável a longo prazo	difícil	fácil	51	3.29	1.04	4	1
gastos com manutenção	fácil	fácil	47	3.43	0.74	4	1
passivo	fácil	fácil	51	3.49	0.92	4	1
ativo	fácil	fácil	51	3.71	0.61	4	2
receitas	médio	fácil	51	3.71	0.58	4	2
despesas	fácil	fácil	51	3.78	0.46	4	2

Fonte: elaborado pelo autor

ANEXO 1 – Ofício de Aprovação do Projeto de Pesquisa pelo CEP

Universidade de São Paulo
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Comitê de Ética em Pesquisa

Campus de Ribeirão Preto

Of.CEtP/FFCLRP-USP/114-dgfs.

Ribeirão Preto, 28 de novembro de 2018.

Prezado(a) Pesquisador(a),

Comunicamos a V. Sa. que o projeto de pesquisa intitulado **“Atenção visual e compreensão de relatórios do setor público: uma abordagem experimental para acessar o entendimento dos millenials.”** foi analisado pelo Comitê de Ética em Pesquisa da FFCLRP-USP, em sua 186ª Reunião Ordinária, realizada em 22.11.2018, e enquadrado na categoria: **APROVADO** (CAAE nº 95178518.3.0000.5407).

Solicitamos que eventuais modificações ou emendas ao projeto de pesquisa sejam apresentadas ao CEP, de forma sucinta, identificando a parte do projeto a ser modificada e suas justificativas. De acordo com a Resolução nº466 de 12/12/2012, devem ser entregues relatórios semestrais e, ao término do estudo, um relatório final sempre via Plataforma Brasil.

Atenciosamente,

Prof.ª Dr.ª Patrícia Nicolucci
Coordenadora

Ao(À) Senhor(a)
Luiza de Miranda Cabral
Programa de Pós-Graduação em Controladoria e Contabilidade da FEARP- USP.

CEP - Comitê de Ética em Pesquisa da FFCLRP USP
Fone: (16) 3315-4811 Avenida Bandeirantes, 3900 - bloco 01 da Administração - sala 07
14040-901 - Ribeirão Preto - SP - Brasil
Homepage: <http://www.ffclrp.usp.br> - e-mail: coetp@ffclrp.usp.br