INTERVENCIÓN COMPORTAMENTAL PARA QUE LAS PERSONAS DE UNA EMPRESA DEL SECTOR CONSTRUCTOR EN MEDELLÍN APRENDAN A IDENTIFICAR ALERTAS DE PHISHING

Behavioral intervention for employees of a construction company in Medellín to learn to identify phishing alerts

MARIA ALEJANDRA ARCILA OSSA
ANDREA BRAVO GIRALDO
SEBASTIÁN VÉLEZ RUIZ
Asesor, docente
Laura Cardona Mesa

UNIVERSIDAD EAFIT
ESCUELA DE HUMANIDADES
MAESTRÍA EN ESTUDIOS DEL COMPORTAMIENTO
MEDELLÍN
2024

Contenido

Lista de tablas

Lista de figuras

Lista de anexos

Resumen

El crecimiento exponencial de la adopción digital en la vida cotidiana de los colombianos nos facilitó el día a día, pero también nos expuso a nuevos riesgos cibernéticos, por ejemplo, el *phishing*, que es un tipo de fraude que consiste en engañar a las personas para que realicen acciones que los expongan a ellos mismos o a sus organizaciones para robar información confidencial o dinero.

Los ataques de phishing no son algo que se evita exclusivamente con tecnología, ya que los atacantes se aprovechan de las personas y de sus sesgos cognitivos. En este trabajo se describe una intervención comportamental tipo Boost en el marco de la maestría de estudios del comportamiento de la universidad EAFIT, que busca mediante un diseño preexperimental que los integrantes de una empresa constructora de la ciudad de Medellín tengan la capacidad de identificar correctamente si un mensaje electrónico tiene alertas de phishing. Abordaremos la conducta objetivo seleccionada, los actores involucrados, el foco comportamental que queremos intervenir, así como los detalles de la intervención, limitaciones, resultados y conclusiones.

Si bien los resultados de la intervención no lograron generar un impacto estadísticamente significativo, en el análisis descriptivo observamos una mejora del 5% en la capacidad de identificar señales de alerta, tanto en mensajes de phishing como mensajes legítimos de instituciones.

Esperamos que este trabajo contribuya con las estrategias de las instituciones y de las personas para hacer frente a este problema.

Palabras clave: Ciencias del comportamiento, prevención de phishing, boost, sesgos cognitivos, SMS, correo electrónico.

Abstract

The exponential growth in digital adoption in the daily lives of Colombians has facilitated our everyday activities but also exposed us to new cybersecurity risks, such as phishing. Phishing is a type of fraud that deceives individuals into taking actions that expose themselves or their organizations and steal confidential information or money.

Phishing attacks cannot be prevented solely through technology, as attackers exploit individuals and their cognitive biases. This paper describes a behavioral Boost intervention within the context of the Master's degree in Behavioral Studies at EAFIT University. It aims, through a pre-experimental design, for members of a construction company in Medellín to be able to correctly identify whether an electronic message contains phishing alerts. We will address the selected target behavior, the individuals and institutions involved, the behavioral focus we aim to modify, as well as the details of the intervention, limitations, results, and conclusions.

Although the intervention did not achieve a statistically significant impact, the descriptive analysis showed a 5% improvement in the ability to identify warning signals, in both phishing messages and legitimate messages from institutions. We hope this work contributes to the strategies of institutions and individuals in addressing this issue.

Keywords: Behavioral sciences, phishing prevention, boost, cognitive biases, SMS, email.

Introducción

La pandemia forzó a las personas a acercarse, conocer, entender y adaptarse al uso de las tecnologías en su vida cotidiana, actividades que anteriormente se realizaban de forma física migraron a la virtualidad facilitando y agilizando su día a día; este suceso logró una aceleración digital en sectores como el financiero, el comercio electrónico, entre otros. Según el informe de gestión del 2021 del grupo Bancolombia, el 85% de las transacciones se realizaron por canales digitales con un aumento del 32% respecto al 2020 (Grupo Bancolombia, 2024) y en el 2023 el incremento fue del 19% respecto al año anterior (Grupo Bancolombia, 2024).

Por otro lado, debido a las condiciones de la pandemia, hubo un crecimiento considerable en las cuentas de régimen simplificado (Nequi, Daviplata, Ahorro a la mano) bancarizando digitalmente personas que no lo estaban (Colombia Fintech, 2022).

No solo el sector financiero tuvo este crecimiento exponencial, un informe de la Cámara Colombiana del comercio electrónico nos muestra que el comercio electrónico en el país tuvo un crecimiento del 84.5% entre el tercer trimestre del 2021 respecto al del 2020 (CCCE, 2021), también el mismo informe destaca que en el primer trimestre del 2023 hubo un aumento del 24.1% respecto al primer trimestre del 2022 y de un 83.2% respecto al mismo trimestre del 2021.

Así como ha crecido el comercio electrónico, se ha incrementado el uso del marketing digital a través de medios como redes sociales, correo electrónico y mensajes de texto, siendo este último uno de los preferidos ya que han emergido como una herramienta de comunicación excepcionalmente eficaz con una tasa de apertura del 90% en los primeros 5 minutos del envío (Blue Radio, 2023).

Los anteriores datos, nos revelan el crecimiento que ha tenido la adopción digital en la vida cotidiana de los colombianos, que, aunque nos facilita nuestro día a día, también nos expone a riesgos cibernéticos que usualmente comienzan como *phishing* y tienen como fin el robo de dinero o de información de las personas y empresas.

El *phishing* es un tipo de fraude que consiste en engañar a las personas para que realicen acciones que los expongan a ellos mismos o a sus organizaciones al ciberdelito, por ejemplo, el robo de información confidencial o de dinero (Greitzer et al., 2021).

Usualmente los mensajes de *phishing* se caracterizan por ser similares al de una entidad confiable (usando sus logotipos y vocabulario), cuentan con un argumento persuasivo (descuentos, ofertas o regalos) o que inciten un sentido de urgencia en la víctima (confirmar una transacción, notificaciones de embargo, entre otros) (Greitzer et al., 2021).

Al analizar el comportamiento del cibercrimen en Colombia encontramos que:

1. El fraude digital de servicios financieros tuvo el mayor aumento porcentual en su tasa, la cual aumentó un 198% durante el periodo de 2019 a 2021 (Transunion, 2023).

2. Durante el 2022, las denuncias en Colombia por ciberdelitos crecieron un 26%, cada 8 minutos se registró una nueva denuncia en el país, siendo el hurto por medios informativos el delito más frecuente: un total de 25.413 equivalente a un 34% más que en el 2021 (Cámara Colombiana de Informática y Telecomunicaciones [CCIT], 2022).

El impacto del *phishing* en la sociedad es significativo y se manifiesta a través de diversos efectos que afectan tanto a las personas como a las instituciones financieras y en última instancia, a la economía en general. Según los informes de gestión anuales de los principales bancos de Colombia, como Davivienda y Bancolombia, se revela que durante el 2022 se enfrentaron a pérdidas económicas considerables asociadas al fraude externo, las cuales corresponden al 89.6% y el 56.2% respectivamente (Grupo Bancolombia, 2023; Banco Davivienda, 2024) Además, las consecuencias de este tipo de ataques incluyen datos

comprometidos, vergüenza social, reducción de confianza interpersonal, en las instituciones y pérdidas económicas (Kelley et al.,2012).

De acuerdo con el estudio realizado por transunion, el miedo por la seguridad y a ser víctima de un fraude puede afectar los ingresos de las empresas, por ejemplo, el 63% de los compradores no volvería a una web si cree que puede ser víctima de un fraude, el 55% admite haberse echado atrás a la hora de abrir una cuenta desde el móvil debido al miedo por la seguridad y el 48% de los compradores abandonó su carrito debido a las preocupaciones relacionadas con la seguridad (TransUnion, 2023).

El *phishing* continuará evolucionando gracias a la utilización de técnicas de inteligencia artificial y automatización, que permitirán a los defraudadores realizar ataques dirigidos y masivos, con mensajes más convincentes que facilite el engañar a los usuarios, incluso se podría suplantar alguna identidad a partir de imágenes, videos y notas de voz (Linkedin Cybertrust Latam, 2024; Tarlogic, 2023).

Por otro lado, su crecimiento acelerado no es un fenómeno exclusivamente colombiano, sino que se observa a nivel mundial. Según el Anti-*Phishing* Working Group (APWG) en el año 2023 se observó casi 5 millones de ataques, siendo este el peor año registrado en materia de *phishing* (Anti-Phishing Working Group, Inc [APWG], 2023).

Es importante anotar que el *Phishing* no es algo que se resuelve exclusivamente con la tecnología, ya que los ciberdelincuentes se aprovechan de las personas y de sus sesgos cognitivos, haciéndolos actuar de manera apresurada e incauta, dado que se hacen pasar por figuras de confianza como empresas o entidades reconocidas cayendo con facilidad y entregando información confidencial. Al analizar el comportamiento de este tipo de fraude en Colombia, se evidencia la necesidad de abordar el problema a nivel comportamental. Este tipo de ataques no solo representan una amenaza para los individuos, sino también para las instituciones financieras, las empresas y la economía en general. La falta de conciencia y

educación sobre los riesgos del phishing contribuye a su persistencia y al éxito de los ataques, haciendo que sea crucial desarrollar estrategias diferentes para prevenirlos y generar capacidades para que reconozcan mensajes fraudulentos y minimizar sus impactos.

Este trabajo busca analizar desde una perspectiva comportamental diferentes técnicas para que las personas del equipo administrativo de una empresa de alquiler de equipo para el sector constructor en la ciudad de Medellín llamada Conequipos, logren identificar de mejor manera señales de Phishing en mensajes de texto y correos electrónicos. En primer lugar este documento aborda el marco conceptual y el planteamiento del problema, en segundo lugar se explicará el diseño de la intervención comportamental para finalizar con las conclusiones y los resultados de la misma.

Marco Conceptual

Los mensajes de Phishing tienen el objetivo de acudir a la facilidad cognitiva, logrando que resulte familiar, que parezca real y que la acción sea fácil de realizar; de esta forma serán procesados por el sistema 1 (rápido, emocional e intuitivo) (Kahneman, 2013).

Dado que el procesamiento del sistema 2 (lento, racional y lógico), exige tiempo y recursos cognitivos, éste se suprime si el destinatario del mensaje está distraído o presionado por el tiempo o el temor. En caso que sea procesado por el sistema 2, se pretende que se realice una evaluación incorrecta de la validez del mensaje (Luo et al., 2013). Al llegar a este punto, se requieren habilidades y conocimientos sobre cómo evaluar la veracidad del contenido, por ejemplo: URL (Particularmente el dominio del link, por ejemplo google.com, netflix.com, etc. pues con frecuencia se utilizan dominios levemente alterados), remitente (es común que el atacante busca imitar a una persona o empresa de confianza) (Weaver y Braly, 2021), alertas generadas por el lector de mensajería (gmail, outlook, Whatsapp, etc.), ortografía, políticas de comunicación de la compañía, entre otros. Estas validaciones

usualmente no son conocidas por las personas o se tienden a olvidar. Existe evidencia de oportunidades de mejora en la manera en que se proporcionan los entrenamientos para estas validaciones, bien sea en términos de formato, interactividad, o de creación de atajos que requieran una menor carga cognitiva (Alkhazi et al., 2022).

De acuerdo a algunos estudios consultados, las siguientes son características de individuos que los hacen más propensos a ser víctimas de phishing:

1. **Comportamiento de riesgo:** Se considera que la forma como asumen el riesgo las personas es una de las principales causas de éxito del ataque de phishing, al igual que pasa con las estafas en el mundo físico. A medida que la posible víctima asuma más riesgos puede caer más fácil en ataques de phishing (Abroshan et al., 2021).
2. **Estilo en la toma de decisiones**: Los estafadores ponen a la víctima en una situación en la que es más probable que tome decisiones con su sistema automático y un error en la toma de decisiones puede aumentar el éxito de una estafa (Abroshan et al., 2021).
3. **Falta de entendimiento del phishing y cómo opera:** La ausencia de entendimiento del phishing y las formas en que opera expone mucho más a las personas a ser víctimas de él (Zhuo et al., 2024).
4. **Falta de conciencia de ser una posible víctima y el impacto que puede traer:** El sesgo de optimismo frente al phishing aumenta la susceptibilidad de caer en éste (Lei et al., 2023). Ser consciente de la vulnerabilidad propia no solo reduce el riesgo de ser víctima, sino que también habilita dos sesgos y emociones que brindan una protección adicional: La aversión al arrepentimiento (Zeelenberg y Pieters, 2007), y el miedo de caer en phishing. Estos aspectos pueden ser de utilidad para subir la prevención de las personas, al capitalizar el miedo y por ende

ser más conscientes a la hora de evaluar un mensaje (Wall y Buche, 2017; Ogbanufe y Pavur, 2023).

Otras características humanas han sido evaluadas en diferentes estudios pero no se lograron resultados concluyentes que indiquen una correlación con ser víctima de phishing, pero de igual manera sugerimos tenerlas en cuenta en potenciales estudios futuros (Por ejemplo, variables extrañas a controlar). Entre estos aspectos se encuentran la edad, el género, la personalidad (Greitzer et al., 2021) o la adicción a redes sociales (Ley et al., 2023).

Revisión y Análisis de Intervenciones

Debido a la heterogeneidad de aspectos que influyen en la probabilidad de ser víctima de phishing, encontramos intervenciones que abordan diferentes características. Por ejemplo, existen intervenciones orientadas a un mejor uso del sistema 1 y el sistema 2 a la hora de evaluar un mensaje de phishing como una intervención basada en mindfulness que logró que los participantes evaluaran mejor los mensajes fraudulentos (Jensen et al., 2017).

También se han llevado a cabo intervenciones orientadas a mejorar el impacto de los entrenamientos en phishing a través de cambios de formato de estos, por ejemplo utilizando formatos visuales en los entrenamientos y comparando su efectividad con entrenamientos solo basados en texto (Jensen et al., 2017), o utilizando juegos (Alkhazi et al., 2022). La inoculación gamificada también se considera una forma efectiva de desarrollar conciencia de la vulnerabilidad de las personas, como se observa en intervenciones para la correcta identificación de noticias falsas en Colombia (Ramírez et al., 2022). También identificamos intervenciones donde se realizan entrenamientos con una retroalimentación detallada e inmediata, con el fin de generar mayor persistencia y conciencia de los aprendizajes (Singh et al., 2023).

Para que la capacidad de identificar mensajes fraudulentos perdure en el tiempo, se ha comprobado la efectividad de enviar recordatorios de las técnicas aprendidas a través de mensajes cortos, como por ejemplo en la intervención de fraude financiero de Jeremy Burke Can educational interventions reduce susceptibility to financial fraud? (Burke, 2022). Para finalizar, existen herramientas y mejores prácticas de entrenamientos de heurísticas que se recomienda usar a la hora de realizar boosts para mejorar la efectividad de identificar mensajes de phishing, cómo enumeran (Weaver y Braly, 2021,).

Alkhazi et al. (2022) identificó que combinar diferentes métodos, como charlas magistrales, juegos e imágenes en la misma intervención, resultó en una mejor actitud de los participantes a la hora de evaluar satisfacción frente al entrenamiento, y a su vez demostró que esta satisfacción influye positivamente en la predisposición a atender futuros entrenamientos sobre Phishing. Esto resulta relevante para nuestra intervención, donde buscaremos mezclar varios de los métodos de la revisión aquí descrita.

Planteamiento del Problema

Descripción del Problema Comportamental

El phishing emplea trucos de ingeniería social (trucos psicológicos que sacan ventaja de la vulnerabilidad y la confianza humana para engañar a sus víctimas y conseguir que voluntariamente ejecuten una acción aprovechable por el criminal para efectuar el ataque) (Thomas, 2006). Los ataques más efectivos de Phishing utilizan los seis principios psicológicos de persuasión propuestos por Cialdini (autoridad, consistencia, simpatía, reciprocidad, escasez y prueba social) (Zhuo et al., 2024). Dado que la acción crucial para prevenirla está del lado de la persona atacada, la conducta que queremos abordar en la víctima de phishing es omitir la acción sugerida por el atacante, ya que identifica que el mensaje contiene alertas que lo categorizan como potencial Phishing. Poder evaluar

heurísticas en un mensaje es una manera efectiva de protegerse contra una gran cantidad de ataques de phishing, como se explicará en la sección de intervenciones analizadas.

Existen múltiples formas de diseñar un ataque de Phishing, por diferentes medios y con diferentes estrategias. El correo electrónico, los mensajes de texto o los mensajes en redes sociales son mecanismos muy comunes que utilizan los ciberdelincuentes. Debido a esto, la conducta puede ocurrir en cualquier lugar físico, mediante un dispositivo electrónico que pueda recibir mensajes en alguna plataforma como las mencionadas.

Existen dos actores principales y varios actores secundarios involucrados en la conducta: En primer lugar está la víctima del ciberdelito, que recibe el mensaje y debe decidir entre entrar al link y proporcionar información o no hacerlo. Está también el atacante que diseña el delito y está atento a las víctimas que caigan en él. Se cuenta además con un tercer actor que es la institución asociada (banco, red social, etc) en cuyo nombre se diseña el ataque. Este actor si bien es pasivo en el ataque en sí, puede sufrir de pérdida reputacional y adicionalmente es un actor fundamental en el reporte y atención a la víctima del ataque.

Adicionalmente identificamos otros actores secundarios involucrados encargados de habilitar la tecnología que permite el ataque, como lo son las empresas de telecomunicaciones (proveedores de mensajes de texto) o plataformas de correo electrónico (Gmail, Outlook, etc.), las empresas que proveen los sistemas operativos de los celulares (Apple, Google, etc.) o computadoras (Dell, Lenovo, etc.), y también los proveedores de envío de comunicaciones masivas, bien sea a través de emails, SMS, Whatsapp, etc.

Para nuestra intervención, nos enfocaremos en los empleados de Conequipos S.A.S, una empresa del sector constructor ubicada en la ciudad de Medellín, Antioquia. Elegimos esta compañía por dos razones: La facilidad de acceso a la empresa y el interés de la compañía para desarrollar mecanismos de reducción de riesgo de phishing como parte de su estrategia de bienestar y seguridad hacia sus empleados.

Nos enfocaremos en los empleados que en su día a día utilizan medios electrónicos para su trabajo y su vida personal, por lo que solamente participarán las personas que hacen parte del área administrativa de la empresa, para un total de 20 personas. El objetivo de la intervención busca impactar de manera personal a los participantes, es decir que las herramientas que recibirán estarán encaminadas a la seguridad personal de sus datos y su dinero.

De acuerdo a los datos recolectados en Conequipos, encontramos que:

- El 100% de las personas utiliza en su día a día el celular y el 75% de las personas usan computadora.
- El 12.5% tienen entre 18 años a 24 años, el 37,5% está entre 25 años y 34 años, 25% tienen entre 35 años a 44 años y 25% mayores a 45 años.
- El 50% de las personas tienen estudios técnicos o tecnológicos, 37,6% tienen estudios de pregrado y el 12,5% son bachilleres.

Para nuestra intervención en Conequipos, luego de analizar impacto, probabilidad de cambio, efecto derrame, y facilidad de medición, concluimos que la conducta objetivo en la que nos enfocaremos es: La víctima abre el mensaje pero identifica rasgos que le generan desconfianza (URL, número de teléfono del remitente, entre otros), y no ejecuta la acción solicitada.

El comportamiento debe ser realizado por la potencial víctima, en el momento en que reciba el mensaje, independiente del espacio físico o la frecuencia en que ocurra. No es necesaria la participación de más personas.

Al realizar un revisión exhaustiva de diferentes intervenciones previas e investigaciones sobre cómo lograr el comportamiento deseado, se identificaron los siguientes aspectos relevantes en capacidad, oportunidad y motivación (COM-B) (Michie et al., 2014).

Capacidad Psicológica

Las capacidades psicológicas juegan un rol clave en la literatura y la investigación alrededor del phishing. En primer lugar, es muy importante contar con un conocimiento base de qué es phishing, cómo identificarlo y el riesgo que conlleva este ciberdelito. Estudios previos mencionan la alta correlación que existe entre el conocimiento previo y la susceptibilidad de ser víctima (Zhuo et al., 2024). Adicionalmente las habilidades de análisis crítico de los mensajes (utilizando el sistema 2) o el desarrollo de atajos para identificar con facilidad mensajes fraudulentos pueden contribuir a una intervención efectiva para lograr la conducta deseada (Zhuo et al., 2024).

Gestionar de manera efectiva el esfuerzo cognitivo que se debe realizar al evaluar un mensaje (Greitzer et al., 2021), y asumir una postura de prevención de riesgos (Abroshan et al., 2021), también son capacidades psicológicas prometedoras.

De acuerdo a nuestro diagnóstico empírico en Conequipos, identificamos que el 37.5% de las personas desconocen el concepto Phishing. y al indagar sobre la percepción de vulnerabilidad frente a este tipo de ataques, los resultados revelaron que el 44% de los encuestados se siente altamente vulnerable. Por otro lado, el 31% indicó sentirse moderadamente vulnerable, mientras que el 25% restante no percibe vulnerabilidad alguna ante dichos ataques. Lo que muestra que las personas perciben esto como un problema.

Motivación Reflexiva

Encontramos dos motivaciones reflexivas relacionadas con la conducta: En primer lugar, el deseo de cuidarse de ataques de phishing, al reconocerse como una posible víctima. Esto contribuye directamente a reducir el sesgo de optimismo, que está correlacionado con una mayor susceptibilidad de ser víctima de phishing (Ley et al., 2023). En segundo lugar

está la aversión al arrepentimiento, que se puede capitalizar para generar mayor precaución con el fin de evitar ser una víctima (Zeelenberg y Pieters, 2007).

Al indagar en Conequipos sobre el nivel de conciencia de ser vulnerable a un ataque de phishing, encontramos que el 43.75% de las personas se consideran probables víctimas de éste en los próximos 12 meses. Y el 75% de las personas manifiestan estar preocupados por su seguridad en línea. Esto sugiere un nivel de conciencia inicial razonable en Conequipos frente a los riesgos del phishing, que es alentador pues la literatura demuestra que quienes no son conscientes de este riesgo son más propensos a caer en estos ciberdelitos (Sykosch et al., 2022).

Motivación Automática

Así como existen motivaciones reflexivas, contamos con dos automáticas que se pueden capitalizar: El miedo ante las amenazas, utilizando la teoría de motivación de protección PMT (Rogers, 1975), es una motivación relevante para las intervenciones alrededor del phishing. Los estudios han demostrado que las apelaciones al miedo son eficaces para promover comportamientos de precaución (Wall y Buche, 2017). Así mismo se puede utilizar la satisfacción de sentirse seguro y protegido, como forma de incentivar prácticas seguras en la vida digital de las personas (Wall y Buche, 2017).

Oportunidad física

Existen diversas oportunidades físicas, entre las que encontramos entrenamiento y educación para que las personas adquieran habilidades de detección de phishing. También las plataformas tecnológicas pueden ayudar con la problemática resaltando de manera más explícita elementos sospechosos en mensajes electrónicos (como remitentes desconocidos o reportados).

Oportunidad psicológica

Una gran oportunidad psicológica se da alrededor de la creación de normas sociales para que las personas reporten y comuniquen de una manera abierta cuando fueron víctimas de phishing. A través de encuestas cualitativas vimos que varias personas han sido víctimas de phishing no hacen público este tema. Ser abiertos a comunicar tanto los peligros como los impactos puede ayudar a generar una conciencia colectiva más adecuada para crear una cultura de prevención frente al phishing.

Diseño de la Intervención

Debido al número de personas que van a participar en la intervención, decidimos implementar un diseño pre-experimental, el cual contará con un pre-test,una intervención comportamental, un post-test a las dos semanas, análisis de dilemas éticos y análisis de resultados.

Para esta intervención utilizamos el modelo behaviour change wheel (BCW) (Michie et al., 2014). Nos enfocamos en 3 dimensiones del COM-B (capability, opportunity, motivation and behaviour model) (Michie et al., 2014). Capacidad psicológica, Motivación automática y Motivación reflexiva, pues no solo las consideramos relevantes sino también complementarias para esta intervención. A continuación, especificamos las funciones de intervención recomendadas y las BCT asociadas:

Función de intervención: Persuasión

Dimensiones de COM-B asociadas: Capacidad Psicológica, Motivación Reflexiva

BCTs priorizadas:

2.2 Retroalimentación de la conducta

2.7 Retroalimentación de los resultados de la conducta

5.1 Informar acerca de las consecuencias de la salud

Función de intervención: Educación

Dimensiones de COM-B asociadas: Capacidad Psicológica, Motivación Reflexiva

BCTs priorizadas:

2.2 Retroalimentación de la conducta

2.7 Retroalimentación de los resultados de la conducta

5.1 Informar acerca de las consecuencias de la salud

Función de intervención: Entrenamiento

Dimensiones de COM-B asociadas: Capacidad Psicológica, Motivación Reflexiva

BCTs priorizadas:

2.2 Retroalimentación de la conducta

2.7 Retroalimentación de los resultados de la conducta

4.1 Instrucción sobre cómo llevar a cabo la conducta

Como mencionamos en la revisión de literatura, existen diferentes enfoques prometedores que queremos profundizar en nuestra intervención que están asociadas a estas BCTs. Tomamos referencias de intervenciones pasadas apropiadas para los atributos de los empleados de Conequipos. Desde las BCTs elegidas nos enfocaremos en mecanismos de retroalimentación de la capacidad de identificar ataques de phishing, y en información general de las consecuencias de un ataque de este tipo. Lo abordaremos tanto desde la educación, la persuasión (utilizando técnicas argumentativas enfocadas a los sesgos asociados al phishing que encontramos) y desde el entrenamiento. Desde la BCT 4.1 (Instrucción sobre cómo llevar a cabo la conducta) abordaremos diferentes señales de alerta para la identificación de un mensaje de phishing.

Hipótesis y Variables

Como variable dependiente tenemos la capacidad de identificar correctamente si un mensaje digital tiene alertas de phishing o no (medido en número de respuestas correctas al clasificar un conjunto de mensajes de texto y correos electrónicos). Como variable independiente está el entrenamiento para detectar mensajes de phishing que se dictará a las personas de Conequipos.

Existen múltiples variables extrañas, entre las que se encuentran: Conocimiento previo en ciberseguridad, experiencia previa como víctima de ataques cibernéticos, nivel de escolaridad, edad y estrato social. Estas variables serán analizadas durante la revisión de los resultados, buscando encontrar hallazgos útiles alrededor de la efectividad para identificar alertas de phishing.

La hipótesis de la intervención está centrada en validar la efectividad del entrenamiento:

H_0: Las personas que reciben un entrenamiento en phishing con técnicas argumentativas y retroalimentación detallada identifican de mejor manera alertas de mensajes de phishing.

Diseño de Instrumento

Con el fin de acercar la intervención a situaciones reales de las personas en su vida diaria, recopilamos 30 mensajes reales (entre mensajes de texto y correos electrónicos) donde 15 de estos mensajes son casos reales de phishing y 15 no. Para decidir las páginas web o plataformas tecnológicas de donde obtendríamos los ejemplos tanto para los tests como para la intervención, consultamos a las personas de Conequipos las plataformas con las que más interactúan en su vida digital, y utilizamos las respuestas para garantizar que los ejercicios de

pre test y post test tuvieran relación con estas plataformas. Así, logramos que todo el marco del instrumento sea cercano a la vida cotidiana de nuestro público objetivo.

Ponderando la cantidad de ejemplos de acuerdo a estas respuestas y garantizando tener un número equivalente de ejemplos que fueran y no fueran phishing llegamos a la siguiente distribución:

Tabla 1

Distribución de los mensajes de acuerdo a las plataformas que usan

Correos electrónicos				Mensajes de texto			
Entidad	# con alertas	# sin alertas	Total	Entidad	# con alertas	# sin alertas	Total
Avianca	1	1	1	Avianca	1	1	2
Bancolombia	1	1	2	Bancolombia	1	2	3
Envíos	1	1	1	Envíos	1	1	2
Facebook	1	1	3	Facebook	1	1	2
Instagram	1	1	2	Instagram	1	2	3
Foto multas	0	0	1	Foto multas	2	0	2
Youtube	0	0	1	Youtube	2	1	3
Netflix	1	1	1	Netflix	0	1	1
Total	6	6	12	Total	9	9	18

*Nota***.** Elaboración propia de acuerdo a las respuestas obtenidas de las personas

Dado que son mensajes reales tanto de phishing como legítimos de las organizaciones, analizamos cada uno para confirmar que en efecto el mensaje solo tiene alertas de phishing si en efecto se tratara de un ciberdelito, o por el contrario, que no existen alertas cuando era un mensaje oficial de la organización o plataforma. Al realizar este análisis encontramos dos tipos de inconsistencias:

1. Mensajes legítimos de organizaciones pero que contienen alertas de phishing como mala ortografía, remitentes sospechosos o estructuras de links que son comunes en ataques cibernéticos.
2. Mensajes de phishing sin alertas evidentes, muy bien elaborados por parte del atacante y utilizando plataformas de envío de mensajería confiables.

Para el pre test y el post test descartamos los mensajes que tuvieran estas inconsistencias, reduciendo el número de preguntas a 24. Ésta es una conclusión a resaltar de nuestro trabajo: Muchas organizaciones envían mensajes legítimos con alertas comunes de mensajes de phishing, y al mismo tiempo, muchos ataques de phishing son muy sofisticados y no cuentan con alertas evidentes. Es necesario entonces, como abordaremos en la intervención, enseñar tips adicionales incluso cuando el mensaje es aparentemente legítimo.

Con el fin de calibrar la dificultad de las preguntas entre pre test y post test, un grupo de 5 personas externas a Conequipos pero con un perfil socio-demográfico similar dieron respuesta a las 24 preguntas. La siguiente tabla sumariza el número de personas que respondieron acertada y equivocadamente cada una:

Tabla 2

Calibración de las preguntas para el pre-test y post-test

ID Pregunta	¿Contiene alertas de phishing?	# de respuestas correctas	% de respuestas correctas	Tipo de mensaje	Entidad
1	No	2	40%	Correo	Avianca
2	No	1	20%	Correo	Bancolombia
3	No	2	40%	Correo	Amazon
4	No	0	0%	Correo	Facebook
5	Si	3	60%	Correo	Simit
6	Si	5	100%	Correo	Instagram
7	Si	5	100%	SMS	Netflix
8	No	4	80%	Correo	Netflix

ID Pregunta	¿Contiene alertas de phishing?	# de respuestas correctas	% de respuestas correctas	Tipo de mensaje	Entidad
9	No	1	20%	SMS	Lifemiles
10	No	5	100%	SMS	Bancolombia
11	Si	5	100%	SMS	Envíos
12	Si	5	100%	Correo	Facebook
13	No	2	40%	SMS	Simit
14	Si	5	100%	SMS	Bancolombia
15	No	2	40%	SMS	Interrapidisimo
16	Si	5	100%	Correo	Facebook
17	No	3	60%	SMS	Secretaría de movilidad
18	Si	5	100%	SMS	Instagram
19	Si	5	100%	SMS	Bancolombia
20	No	0	0%	SMS	Facebook
21	Si	5	100%	SMS	Instagram
22	Si	5	100%	SMS	Facebook
23	Si	5	100%	Correo	Bancolombia
24	No	3	60%	Correo	American Eagle

Nota. Construcción propia con las respuestas del test de calibración

Con base a estos resultados creamos los siguientes tests que están equilibrados de acuerdo a su nivel de complejidad y que incluye, cada uno, 6 preguntas con alertas de phishing y 6 preguntas que no contienen ninguna:

Tabla 3

Distribución de los mensajes del pre-test y el post-test

Pre-test				Post-test			
ID Pregunta	Tipo de mensaje	Entidad	Contiene alertas	ID Pregunta	Tipo de mensaje	Entidad	Contiene alertas
1	Correo	Avianca	No	3	Correo	Amazon	No
2	Correo	Bancolombia	No	4	Correo	Facebook	No

6	Correo	Instagram	Si	5	Correo	Simit	Si
7	SMS	Netflix	Si	8	Correo	Netflix	No
10	SMS	Bancolombia	No	9	SMS	Lifemiles	No
13	SMS	Simit	No	11	SMS	Envíos	Si
14	SMS	Bancolombia	Si	12	Correo	Facebook	Si
16	Correo	Facebook	Si	15	SMS	Interrapidisimo	No
19	SMS	Bancolombia	Si	17	SMS	Secretaría movilidad	No
20	SMS	Facebook	No	18	SMS	Instagram	Si
22	SMS	Facebook	Si	21	SMS	Instagram	Si
24	Correo	American Eagle	No	23	Correo	Bancolombia	Si

*Nota***.** Construcción propia de acuerdo a las respuestas obtenidas de las personas.

Consideraciones éticas

Durante el diseño de la intervención se discutieron los siguientes dilemas éticos y la mejor manera de gestionarlos durante la misma. Para futuras intervenciones recomendamos considerarlos.

1. **Uso del miedo durante la intervención:** Para la intervención es necesario generar conciencia del riesgo del phishing, y desde la perspectiva de la motivación automática queremos utilizar la PMT (Teoría de motivación a la protección), generando cierto nivel de precaución de los empleados frente a su vida digital. Es importante mantener un discurso balanceado entre el miedo y la seguridad que proporcionan las técnicas que se enseñan, para evitar una reacción emocional poco efectiva y evitar generar parálisis y desconfianza generalizada.

2. **Uso de información personal de los empleados:** Los ataques de phishing más efectivos tienen cierto nivel de personalización y de conocimiento previo de las posibles víctimas, para así aumentar la probabilidad de caer en éstos. Los instrumentos de diagnóstico con frecuencia utilizan datos personales de cada sujeto, lo que podría conllevar a un uso indebido de información personal. Para garantizar un uso ético de la información se firmó con los participantes un consentimiento informado y un acuerdo de tratamiento de datos con fines académicos donde se les explicó claramente la intervención en la que participarían, la confidencialidad de la información, el uso agrupado de los datos, y que se destruiría la información personal una vez terminado el análisis de este proceso.
3. **Enseñar técnicas de phishing puede capacitar potenciales atacantes:** Se podría argumentar que entrenar a las personas en phishing los capacita para ellos mismos desarrollar nuevos ataques. Si bien es claro que el conocimiento de las técnicas abre esta puerta, priorizamos la capacidad tan necesaria que tienen las personas de defenderse en el mundo actual.
4. **Generar desconfianza hacia las instituciones que se utilizan en los ejemplos y las pruebas:** Nuestra metodología usó ejemplos reales de ataques de phishing de entidades cercanas al público objetivo. Esto podría potencialmente generar una desconfianza a estas instituciones en particular. Para mitigar este riesgo, durante el entrenamiento se habló de una manera clara que el phishing es una problemática general y garantizamos que también existen ejemplos legítimos de mensajes de cada una de estas instituciones, tanto en la intervención como en los tests.

Desarrollo de la Intervención

Inicialmente se realizó el pretest de manera presencial con los 17 empleados administrativos de Conequipos que estaban disponibles durante las fechas de la intervención. El 21 de Marzo a las 9 AM nos encontramos con las 17 personas en la sede de Conequipos. En primer lugar, cada participante diligenció un formulario físico con el consentimiento informado (Anexo 5) y una autorización al tratamiento de datos personales (Anexo 6), donde se explicó que eran parte de una intervención con fines académicos.

Posteriormente cada persona diligenció un formulario digital con las 12 preguntas del pretest, así como diferentes preguntas demográficas. A las personas se les aclaró que en el pretest y en el post test existían ejercicios tanto que eran phishing como otros que no lo eran, con el objetivo de evitar que las personas estuvieran muy prevenidas.

Durante la misma sesión, posterior al pretest, se realizó una capacitación con todos los empleados. La capacitación fue realizada por una persona que tuviera la habilidad de hablar con un lenguaje cercano al público de Conequipos, y que utilizó diferentes técnicas argumentativas para facilitar la interiorización de los conceptos. La capacitación duró una hora y contó con las siguientes fases:

1. A través de un juego donde las personas debían identificar que un mensaje tenía alertas de phishing se les mostró su nivel de vulnerabilidad frente a estos ataques. Buscamos a través de este ejercicio hacer muy evidente el sesgo de familiaridad, que es muy utilizado por los atacantes de phishing.
2. Posteriormente (implementando la BCT 5.1 de informar sobre las consecuencias en la salud) se les mostró los riesgos que esto puede traer para sus vidas personales y que tan comunes son estos ataques. Todo esto buscó generar conciencia de la vulnerabilidad de ser víctima buscando generar una sensación de miedo que brinde una protección adicional.

3. Se procedió luego a explicar algunas metáforas y las técnicas más comunes que usan los atacantes para explotar los sesgos de las víctimas de phishing.
4. Se presentó el "ABC del phishing" (Anexo 4) donde se enumeran en un lenguaje cercano diferentes heurísticas para reconocer fácilmente señales de fraude en un mensaje de phishing. En este ABC utilizamos el sesgo de prominencia para resaltar los factores claves a analizar cuando se recibe un mensaje, así como el sesgo del presente, para generar la reflexión de qué sucedería si no se existe reacción ante el mensaje. También se abordó el sesgo de similitud, debido a que los atacantes lo utilizan para que las personas sean más propensas a caer en el problema, por ejemplo, con dominios de correo electrónico o con URLs muy similares a las legítimas.
5. A través de un juego se realizaron varios ejercicios de detección de alertas de phishing, brindando retroalimentación detallada del ejercicio (BCT 2.2).

El día 4 de abril, a las 2 PM, nos reunimos de nuevo de manera presencial con las 17 personas de Conequipos. Allí se diligenció el post test con las 12 preguntas seleccionadas. Adicionalmente en esta sesión se hizo entrega de un infográfico (Anexo 3) con recordatorios simples de las heurísticas de alertas de ataques electrónicos. También se les hizo entrega de la presentación para que la replicarán en el futuro en la organización.

Resultados de la Intervención

Perfil Sociodemográfico

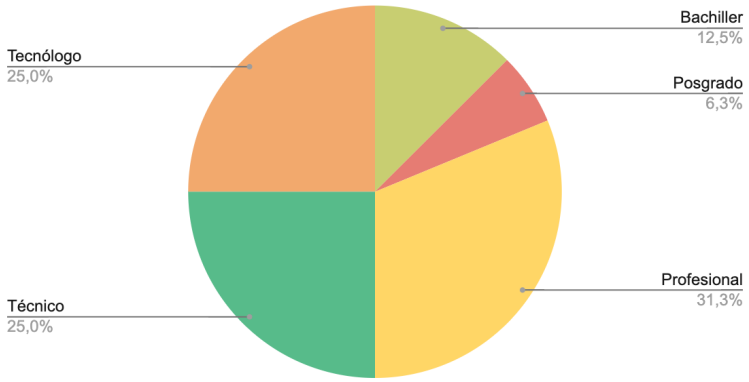
Con base en la información recolectada en el pretest obtuvimos el siguiente perfil sociodemográfico de nuestra población:

El 50% de la población tiene un título técnico o tecnólogo, el 37.6% tienen título universitario o de postgrado y solo el 12.5% es bachiller (ver figura 1), en cuanto a la edad encontramos que la población mayoritaria se encuentra entre 25 y 44 años (figura 2).

La mayoría de las personas manifestaron que no han tenido capacitación en seguridad informática, phishing o fraude (ver figura 6). Sin embargo la mayor parte de los encuestados dicen conocer qué es phishing (figura 4) y declaran haber tenido experiencias cercanas de robo de información o dinero de manera digital (ver figura 5).

Figura 1

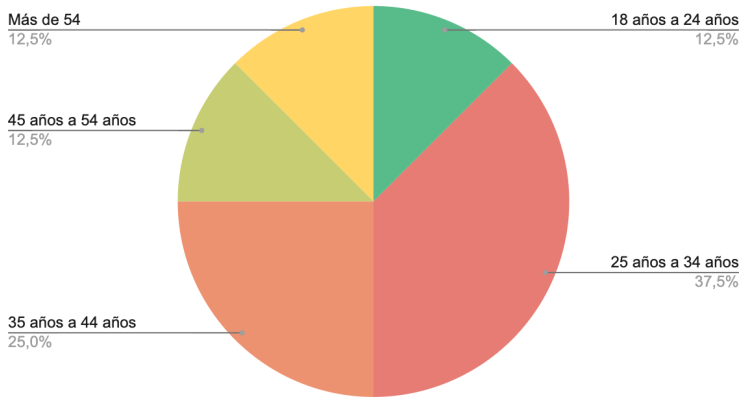
Nivel de escolaridad de la población objetivo



*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Figura 2

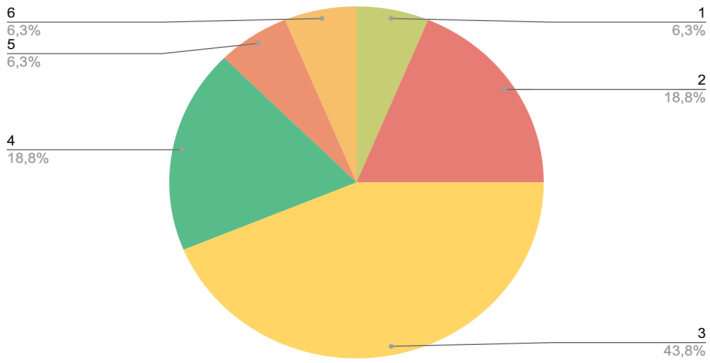
Rangos de edad de la población objetivo



*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Figura 3

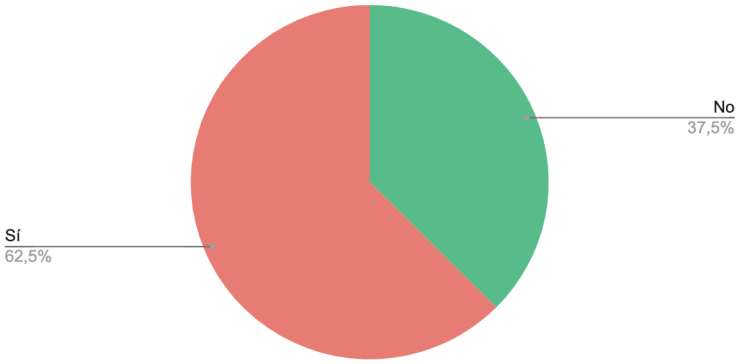
Estrato social de la población objetivo



*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Figura 4

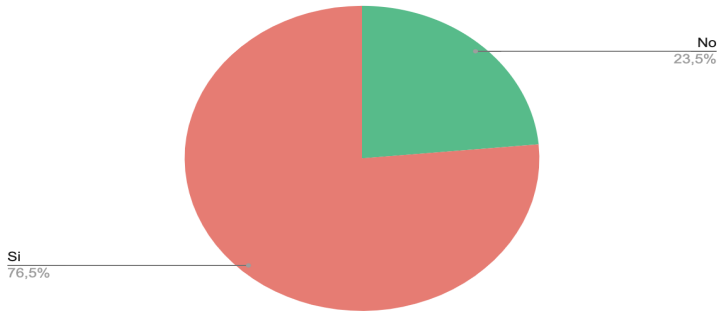
Conocimiento sobre el phishing de la población objetivo



*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Figura 5

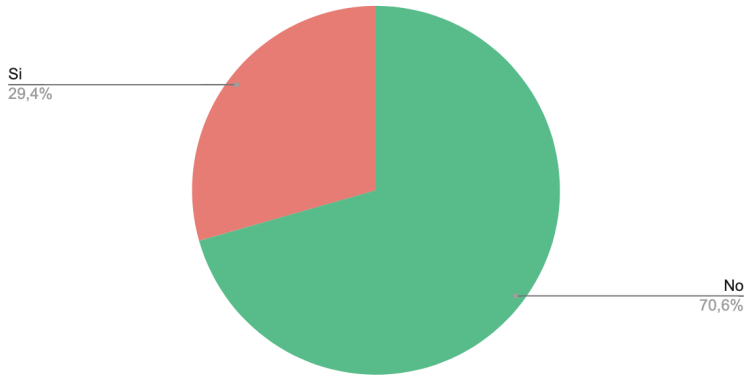
Personas que han tenido experiencia cercana al phishing, a través de robo de dinero o información por medios digitales



*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Figura 6

Personas que han recibido capacitación sobre seguridad informática, phishing o fraude digital en el pasado



*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Resultados Obtenidos

Al consolidar los resultados de las 17 personas intervenidas tanto en pre-test como en post-test, mediante un análisis cuantitativo descriptivo encontramos los siguientes hallazgos:

1. Se observa que algunas personas mejoraron su respuesta versus lo obtenido en el pretest, mientras que otras personas lo empeoraron. Sumando todas las respuestas correctas de toda la población, obtuvimos que se contestaron de manera correcta 10 preguntas más en el post test que en el pre test, mejorando en cerca de 5% el promedio de acierto.

2. Observamos una capacidad mucho más alta de las personas de identificar mensajes que en efecto eran phishing que de identificar correctamente mensajes que no eran phishing. Al distribuir el acierto entre estas dos categorías encontramos que la capacidad de identificar correctamente mensajes que son phishing fue de 87% en promedio en el pre test y 92% en el post test. Para aquellos que no eran phishing tuvimos un 52 y un 57% respectivamente (ver tablas 4 y 5 respectivamente). La mejora pre test - post test fue equilibrada en ambos casos, pero la línea base era muy diferente entre los dos. Ésto, como lo veremos en las limitaciones del ejercicio, creemos que puede estar influenciado por el hecho de estar realizando el ejercicio en un contexto controlado donde se les pidió a las personas que busquen señales de fraude. Sugerimos que para estudios posteriores se realicen pruebas en entornos más naturales y se observe si esta distribución de aciertos cambia.

Tabla 4

Respuestas en número y porcentaje del pre-test y el post-test de los mensajes que sí tienen alertas de phishing

Pre-test				Post-test			
Mensaje	alertas	# ok	% ok	Mensaje	alertas	# ok	% ok
3. Mail Instagram	Si	14	82%	3. Mail Fotomulta	Si	15	88%
4. SMS Netflix	Si	15	88%	6. SMS Envíos	Si	17	100%
7. SMS Bancolombia	Si	16	94%	7. Mail Facebook	Si	17	100%
8. Email Facebook	Si	16	94%	10. SMS Instagram	Si	15	88%
9. SMS Bancolombia	Si	11	65%	11. SMS Instagram	Si	17	100%
11. SMS Facebook	Si	17	100%	12. Mail Bancolombia	Si	13	76%
Promedio			87%	Promedio			92%

*Nota***.** Construcción propia de acuerdo a las respuestas obtenidas de las personas.

Tabla 5

Respuestas en número y porcentaje del pre-test y el post-test de los mensajes que no tienen alertas de phishing

Pre-test				Post-test			
Mensaje	Alertas	# ok	% ok	Mensaje	Alertas	# ok	% ok
1.Mail Avianca	No	10	59%		No	11	65%
2. Mail Bancolombia	No	10	59%		No	8	47%
5. SMS Bancolombia	No	10	59%		No	12	71%
6. SMS Simit	No	7	41%		No	12	71%
10. SMS Facebook	No	2	12%		No	4	24%
12. Email Tigo	No	14	82%		No	11	65%
Promedio			52%				57%

*Nota***.** Construcción propia de acuerdo a las respuestas obtenidas de las personas.

A pesar de esta diferencia en la línea base, la mejora de 5% en el acierto general de la población fue equilibrada entre los mensajes que eran phishing y los que no, por lo que no se observa una diferencia de efectividad en la intervención entre estas dos categorías.

1. Nos llama particularmente la atención, y a su vez nos preocupa, la menor capacidad que tuvieron las personas para identificar correctamente mensajes de phishing que provienen de instituciones financieras. Tanto en el pre test como en el post test, los mensajes de phishing bancarios fueron los que peor capacidad de identificación tuvieron (65% y 76% de acierto en pre test y post test, cuando la media de todo el acierto de mensajes de phishing fue de 87% y 92% respectivamente). Consideramos que esto se debe a lo sofisticado de estos ataques de phishing, que vale la pena repetir, fueron obtenidos de casos reales que actualmente se encuentran en circulación. En el caso del SMS el número de envío era un número corto muy similar a los mensajes legítimos de la institución, y en el

correo electrónico, por su parte, se usaba la imagen y el tono oficial de ésta misma.

2. Por el contrario, las instituciones donde observamos mayor desconfianza de la población fueron las plataformas de envío de paqueteo y Facebook. En estos dos casos observamos que las personas identificaron correctamente los ejercicios que en efecto eran phishing a un 100%, pero los que no eran phishing solo obtuvieron una efectividad de correcta identificación del 24% y el 47% respectivamente, como se observa en la tabla 5 (comparado con un 57% de media de efectividad). Si bien no contamos con una conclusión categórica sobre la razón de estos puntajes, tenemos varias hipótesis: Que se deba a una posible falta de familiaridad con estas entidades, lo que incrementa la desconfianza; que exista una prevención inicial de nuestra población frente a éstas, que no fue detectada en la intervención, o que estas instituciones deben trabajar urgentemente en mejorar la calidad de la comunicación con sus clientes, pues, repetimos, los mensajes que no eran phishing efectivamente fueron mensajes oficiales que encontramos en su comunicación habitual.
3. Durante la intervención presencial una de las personas de Conequipos manifestó una experiencia cercana al phishing relacionada con Facebook. Es posible también que este factor haya generado desconfianza frente a esta institución. Este suceso lo consideramos una variable extraña que no fue posible controlar en la intervención por lo espontánea que fué.
4. Al analizar los resultados obtenidos entre pre test y post test de una manera individual, observamos varios patrones que nos parecen interesantes a pesar del tamaño de la población. En primer lugar, que la persona con el menor nivel socioeconómico fue la persona que peor resultado obtuvo en el pre test pero

también fue la persona que tuvo porcentualmente una mayor mejoría durante la intervención (33% de mejora en efectividad en el post test, como se puede observar en la tabla 6). En futuras intervenciones recomendamos analizar si un perfil socioeconómico más bajo se correlaciona con la ausencia de algunos conceptos base (como falta de acceso a educación financiera o menos familiaridad con comunicación digital) que no fueron identificados en nuestra intervención de manera explícita.

Tabla 6

Respuestas correctas por persona en el pre-test y el port-test

ID Persona	Edad	Estrato	¿Has recibido información previa?	¿ ha sido víctima?	#R correctas Pre-test	%R correctas Pre-test	# R correctas Post-test	% R correctas Post-test	Dif aciertos Pre-test vs Post-test
1	38	6	Si	Si	9	75%	12	100%	25%
2	61	6	No	Si	9	75%	12	100%	25%
3	44	3	No	Si	8	67%	10	83%	17%
4	38	5	Si	Si	11	92%	9	75%	-17%
5	25	3	No	No	9	75%	6	50%	-25%
6	24	2	No	Si	9	75%	10	83%	8%
7	55	3	No	Si	8	67%	7	58%	-8%
8	54	4	Si	Si	9	75%	7	58%	-17%
9	21	3	Si	Si	9	75%	6	50%	-25%
10	27	5	Si	Si	9	75%	11	92%	17%
11	26	3	No	Si	9	75%	11	92%	17%
12	44	3	No	No	7	58%	10	83%	25%
13	31	2	No	No	5	42%	7	58%	17%
14	54	3	No	No	10	83%	10	83%	0%
15	32	3	No	Si	10	83%	9	75%	-8%
16	38	2	No	Si	7	58%	7	58%	0%
17	28	1	No	Si	4	33%	8	67%	33%

*Nota***.** Construcción propia de acuerdo a las respuestas obtenidas de las personas.

1. Si bien los mejores puntajes en el post test vinieron de las 2 personas del estrato socioeconómico más alto (6), no encontramos ninguna otra correlación entre el nivel socioeconómico de los participantes y los resultados del post test.

2. Adicionalmente nos llama la atención que entre las 3 personas con menor edad (Identificadas en la tabla 6 con los IDs 5,6 y 9) se encuentran los 2 resultados con peor efectividad de la intervención, donde la eficacia para identificar mensajes de phishing fue peor en el post test que en el pre test en un 25%. Si bien por lo pequeña de la población no es posible afirmarlo de manera estadísticamente significativa, es posible que nuestra intervención no haya sido efectiva para un público más jóven. Más allá de éste factor, no encontramos ninguna otra conclusión con base en la edad de los participantes de la intervención.
3. Analizando la información vemos que la variable de tener información previa sobre el phishing no contribuyó a un mejor resultado y de hecho los puntajes del post test fueron peores al pre test en los participantes que manifestaron contar con esta información previa. No consideramos que nuestra intervención haya deliberadamente perjudicado a personas con este conocimiento, sino que por el contrario simplemente esta variable no influyó en el desempeño. Quizás otro factor a tener en cuenta es que cuando indagamos un poco más por lo que entendían por entrenamiento en seguridad en línea encontramos algunas respuestas que ellos creían que se trataba de entrenamiento de phishing cuando en realidad no lo era. Sugerimos para futuras intervenciones profundizar tanto en la calidad del contenido como en la antigüedad de estos entrenamientos previos.

Dentro del análisis cuantitativo, en primer lugar realizamos una prueba para determinar la normalidad de los datos tanto de pre test como de post test, usando la prueba de Shapiro-Wilk. El pre-test arrojó un valor estadístico de .857 con un nivel de significancia p=.014, lo cual señala una desviación significativa de la normalidad. Por el contrario, la prueba del post-test mostró un valor estadístico de .919 y un nivel de significancia p = .144,

no presentando evidencia significativa de desviación de la normalidad. como se puede ver en la siguiente tabla.

Tabla 7

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro - Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pretest	0.290	17	<0.001	0.857	17	0.014
Postest	0.185	17	0.124	0.919	17	0.144

a.Lilliefors Significance Correction

*Nota***.** Elaboración propia de acuerdo a las respuestas obtenidas de las personas.

Debido a que una de las muestras es no-paramétrica, procedimos entonces a realizar un test de Wilcoxon para 2 muestras relacionadas no paramétricas. El análisis muestra (ver tabla 8) que si bien se encuentran más participantes con mejora que con detrimento después de la intervención (Positive Ranks vs Negative Ranks), el estadístico entrega un valor p > 0.05 que concluye que estadísticamente no hay evidencia suficiente para concluir que existen diferencias significativas entre las puntuaciones del pre test y el post test.

Tabla 8

Wilcoxon Signed Ranks Test

Ranks				
		N	Mean Rank	Sum of Ranks
Postest - pretest	Negative Ranks	6[a]	6.83	41.00

	Positive Ranks	9[b]	8.78	79.00
	Ties	2[c]		
	Total	17		

a. postest < pretest
b. postest > pretest
c. postest = pretest

Test Statistics[a]

	pretest - postest
Z	-1.092[b]
Asymp.Sig. (2 Tailed)	0.275

a. Wilcoxon Signed Ranks Test.
b. Based on negative ranks.

*Nota***.** Elaboración propia con los resultados estadísticos.

Por último, queremos revisar si haber tenido una experiencia directa o cercana de ser víctima de phishing en el pasado pudo haber contribuido a unos mejores resultados. Al realizar una prueba Shapiro-Wilk de normalidad donde diferenciamos la normalidad de resultados discriminando por esta variable "víctima" se concluye (ver tabla 9) que para el post test siguen una distribución normal ($p > 0.05$) en ambos grupos, los que han sido víctimas de phishing como los que no.

Tabla 9

Tests of Normality cuando las personas han sido víctimas de phishing

		Kolmogorov-Smirnov[a]			Shapiro - Wilk		
	Víctima	Statistic	df	Sig.	Statistic	df	Sig.
Pretest	0	0.214	4	.	0.963	4	0.798
	1	0.302	13	0.002	0.801	13	0.007
Postest	0	0.302	4	.	0.827	4	0.161
	1	0.163	13	0.200	0.930	13	0.344

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

*Nota***.** Elaboración propia con los resultados estadísticos.

Por ende, procedemos a realizar una prueba t-student de muestras independientes para analizar si existe una diferencia estadísticamente significativa en los resultados obtenidos (ver tabla 10). Se obtuvo una significancia mayor de 0.05, por lo que se concluye que no hay una diferencia significativa entre ambos grupos (los que habían sido víctima de phishing en el pasado y los que no), indicando que ser víctima de fraude no tiene un efecto especial sobre la habilidad de detectarlo. Esto en ambos casos, asumiendo o no varianzas iguales.

Tabla 10

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence interval of Difference	
		F	Sig.	t	df	One - Sided p	Two - Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Postest	Equal Variances Assumed	0.08	0.931	-0.775	15	0.0225	0.451	-0.904	1.167	-3.390	1.583
				-0.769	4.957	0.238	0.477	-0.904	1.175	-3.933	2.125

*Nota***.** Elaboración propia con los resultados estadísticos.

Para finalizar nuestro análisis nos quedamos con la percepción si bien los resultados de nuestra intervención no fueron estadísticamente significativos, descriptivamente observamos una mejoría en la eficacia para detectar mensajes de phishing de los participantes. Creemos que el tamaño limitado de la población genera cierto nivel de ruido a la hora del análisis estadístico, por lo que sugerimos para futuras intervenciones replicar los conceptos utilizados en una muestra más grande.

Limitaciones

1. El mecanismo usado de pre-test y post-test consistió en responder si se evidenciaban alertas de phishing o no al ver mensajes en un formulario estático. Esto puede generar predisposición y un comportamiento más racional y conservador a la hora de evaluar mensajes que el que se tiene en la vida cotidiana.

Si bien observamos la mejoría en la evaluación de alertas, no podemos garantizar con este enfoque cómo reaccionan las personas en un contexto más cotidiano. A su vez, esta variable pudo influir en que las personas respondieron muchos más mensajes de los correctos como phishing.

2. La evaluación del impacto de la intervención se dió dos semanas después, limitando el entendimiento de la durabilidad del efecto de éste en el largo plazo.
3. El tamaño de la muestra fue muy corto por las limitaciones de la población. Esto pudo haber afectado el análisis de impacto de la intervención.
4. La manera en la que presentamos los tests a través de imágenes no permite aplicar todas las técnicas enseñadas. Por ejemplo, el uso de herramientas de navegadores web (como Google Chrome) para encontrar alertas de phishing en correos electrónicos o en páginas web.
5. Para controlar variables extrañas se definió que nuestra intervención requería estar presencialmente para garantizar que no se compartiera información entre personas y que no se revisará en paralelo material en internet como apoyo durante los tests. Esto puede afectar la escalabilidad de la intervención.

Conclusiones

Nuestra intervención no logró generar un impacto estadísticamente significativo, y analizando los datos de manera descriptiva, concluimos que la principal razón fue la baja capacidad de detectar mensajes que no eran phishing. El puntaje obtenido promedio en detección correcta de mensajes que no eran phishing fue mucho menor que el puntaje promedio en mensajes que sí eran phishing. Creemos que el ambiente controlado en el que se realizaron las pruebas tiene una incidencia como lo mencionamos en la sección anterior. Sin

embargo, también es posible que las personas estén manifestando un sesgo o una prevención que no se detectó en el estudio.

A su vez, y con la limitación de nuestro tamaño de muestra, no encontramos diferencias significativas estadísticamente al analizar si experiencias previas cercanas al phishing implican una mejor capacidad de identificación de ataques. Si bien encontramos algunos hallazgos descriptivos a nivel de edad y estrato social, no realizamos pruebas estadísticas con estas variables debido a la distribución de datos de nuestra población, pero recomendamos profundizar en este tema en futuras investigaciones asociadas.

En el análisis cuantitativo descriptivo, vimos una mejora del 5% en la capacidad de identificar señales de alerta tanto en mensajes de phishing como mensajes legítimos de instituciones, por lo que podemos interpretar que la intervención pudo tener un impacto si bien no estadísticamente significativo pero que puede ser relevante en poblaciones más grandes.

Es muy importante que las instituciones, tanto públicas como privadas, se enfoquen en enviar comunicaciones de mejor calidad a sus clientes. Al analizar mensajes legítimos para incluirlos en nuestras pruebas nos vimos en la necesidad de descartar una gran cantidad pues a pesar de ser legítimos contenían claras señales de phishing, siendo los más comunes: Errores de ortografía, comunicación poco clara, fuentes de envío poco confiables, y enlaces no verificables. Son errores básicos que pueden generar desconfianza y confusión.

Por el contrario, los atacantes de phishing están implementando mensajes muy sofisticados y sin señales claras de alerta. Esto nos lleva a recomendar para futuros trabajos que no es suficiente enfocarse en las alertas del mensaje como tal, sino también en qué precauciones o heurísticas manejar una vez se descartan todo tipo de alertas, bien sea precauciones al dar click a un enlace, o la validación del mensaje comunicándose por cuenta propia con la entidad asociada. Nuestro entrenamiento incluyó contenido relacionado dentro

de las técnicas para prevenir el phishing, sin embargo, no lo evaluamos dentro del pre test ni el post test.

En el momento que se realizó el post test, varias personas manifestaron que habían implementado en su día a día el hábito de ir a la fuente directa o institución cuando les causa desconfianza un mensaje. Si bien este no era el comportamiento objetivo de la intervención si observamos un beneficio para los empleados de Conequipos pues se generó mayor conciencia y prevención a la hora de recibir un mensaje.

De hecho, este evento plantea la discusión de si la conducta objetivo que elegimos fue la más adecuada. Al ver tanto la sofisticación de los ataques de phishing como la falta de buenas prácticas de comunicación por parte de las instituciones, quizás un mejor enfoque para esta problemática no va tanto en la correcta identificación de alertas en los mensajes digitales, sino más bien en la generación de una conciencia de la vulnerabilidad sumado a simplemente no seguir literalmente las instrucciones del mensaje recibido, sino por el contrario, comunicarse directamente con las instituciones, independiente de las señales de alerta percibidas en el mensaje. Para el contexto de nuestra población y de las instituciones evaluadas esta aproximación nos parece más interesante para futuras intervenciones.

Consideramos que si bien nuestra intervención tuvo resultados positivos y fue bien recibida por el público, cuenta con varias oportunidades de mejora: En primer lugar creemos que fue corta y sin espacios de refuerzo. Quizás el haber enviado el contenido presentado en la intervención posteriormente hubiera sido de ayuda (los mismos empleados de la compañía nos lo pidieron pero no accedimos para controlar que todas las personas tuvieran las mismas condiciones de estudio de cara al post test), o quizás tener varias sesiones o recordatorios paulatinos podría haber ayudado a profundizar de mejor manera los conceptos.

En segundo lugar creemos que futuras intervenciones similares deben contar con un mayor énfasis en cómo identificar mensajes que no eran phishing, pues desde la línea base observamos la dificultad de determinar correctamente si lo eran o no.

En tercer lugar, creemos que existe una oportunidad de aprovechar mejor los espacios de juego dentro de la intervención. En el juego que implementamos, utilizamos ejercicios con imágenes para garantizar una consistencia con la manera como evaluamos en el post test. Creemos que esto pudo haber limitado el efecto del aprendizaje, y quizás haber generado simulacros más cercanos a la realidad, o incluso fomentar juegos de rol cómo "crea tu propio phishing" podría haber permitido una mejor comprensión de los conceptos.

Para futuros trabajos asociados a este tema recomendamos cambiar las condiciones en las que se ejecutan los tests, buscando que sea un ambiente más natural y en medio del día a día de las personas. También invitamos a involucrar el aprendizaje de navegadores web o herramientas tecnológicas en la intervención (por ejemplo la validación de encabezados HTTPS en un navegador como Google Chrome). Por último, para este trabajo descartamos utilizar técnicas como mindfulness o recordatorios frecuentes. Creemos que vale la pena retomar estas técnicas como posibles complementos para intervenciones relacionadas.

Referencias

Abroshan, H., Devos, J., Poels, G., & Laermans, E. (2021, 03 17). Phishing Happens Beyond Technology: The Effects of Human Behaviors and Demographics on Each Step of a Phishing Process. *IEEE Access*, *9*, 44928 - 44949. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9380285

Alkhazi, B., Alshaikh, M., Alkhezi, S., & Labbaci, H. (2022). Assessment of the Impact of Information Security Awareness Training Methods on Knowledge, Attitude, and Behavior. *IEEE*, *10*, 132132-132143. 10.1109/ACCESS.2022.3230286.

Anti-Phishing Working Group, Inc [APWG]. (2023, 11 13). *Phishing Activity Trends Reports*. https://apwg.org/trendsreports/

Banco Davivienda. (2024, March 20). *Informe de gestión 2023*. https://ir.davivienda.com/wp-content/uploads/2024/03/Informe-Anual-Banco-Davivienda-2023.pdf

Bautista García, F., Mesa Guzmán, L., & Blanco, L. F. (2023). *Informe anual de ciberceguridad*. Cámara Colombiana de Informática y Telecomunicaciones. https://www.ccit.org.co/wp-content/uploads/estudio-anual-de-ciberseguridad.pdf

Benenson, Z., Gassmann, F., & Landwirth, R. (2017). Unpacking Spear Phishing Susceptibility. *InInternational conference on financial cryptog-raphy and data security*, *10323*, 610–627. https://doi.org/10.1007/978-3-319-70278-0_39

Blue Radio. (2023, 10 24). *¿Qué tan efectivo es enviar mensajes de texto para aumentar las ventas?* https://www.bluradio.com/tecnologia/que-tan-efectivo-es-enviar-mensajes-de-texto-para-aumentar-las-ventas-pr30

Burke, J., Kieffer, C., Mottola, G., & Perez Arce, F. (2023, 06). Can educational interventions reduce susceptibility to financial fraud? *Journal of Economic Behavior & Organization*, *198*, 250-266. https://doi.org/10.1016/j.jebo.2022.03.028

Cámara Colombiana de Informática y Telecomunicaciones [CCIT]. (2022). *Ciberseguridad en redes de telecomunicaciones moviles*. https://www.ccit.org.co/wp-content/uploads/ciberseguridad-en-redes-de-tel-2022-2.pdf

Cámara Colombiana de Comercio Electrónico. [CCCE]. (2021, 12 2). *Informe de comercio electrónico tercer trimestre de 2021*. https://www.ccce.org.co/gestion_gremial/informe-del-comercio-electronico-en-el-tercer-trimestre-de-2021/

Colombia Fintech. (2022). *Home*. https://colombiafintech.co/

Escobar, J. (2023, 06 01). *Los delitos cibernéticos se han reducido en el 2023: Policía Nacional*. Radio Nacional Colombia. https://www.radionacional.co/actualidad/delitos-ciberneticos-en-colombia-estadisticas-actuales

Greitzer, F. L., Li, W., Laskey, K. B., Lee, J., & Purl, J. (2021, 06 28). *Experimental Investigation of Technical and Human Factors Related to Phishing Susceptibility*. ACM Digital Library.

Grupo Bancolombia. (2017). *10 consejos de Bancolombia para cuidarse de fraudes en temporada de primas*. https://www.bancolombia.com/acerca-de/sala-prensa/noticias/inclusion-educacion-financiera/diez-consejos-para-cuidarse-de-fraudes-en-temporada-de-primas

Grupo Bancolombia. (2024). *Informe de gestión 2023*. https://www.grupobancolombia.com/wcm/connect/5f1aa5f4-4bb0-446d-8030-409e8109c820/Informe_de_gestion_2023.pdf?MOD=AJPERES

Grupo Bancolombia. (s.f.). *Informe de Gestión Grupo Bancolombia 2022*. https://www.grupobancolombia.com/corporativo/informe-gestion

Jansen, J., & Schaik, P. v. V. (2019, 04). The design and evaluation of a theory-based intervention to promote security behaviour against phishing. *International Journal of Human-Computer Studies*, *123*, 40 - 55. https://doi.org/10.1016/j.ijhcs.2018.10.004

Jensen, M. L., Dinger, M., Wright, R. T., & Thatcher, J. B. (2017, 08 17). Training to Mitigate Phishing Attacks Using Mindfulness Techniques. *Journal of Management Information Systems*, *34*(2), 597-626. https://doi.org/10.1080/07421222.2017.1334499

Kahneman, D. (2013). *Pensar rápido, pensar despacio*. Debolsillo.

Kelley, C., Wha Hong, K., Mayhorn, C. B., & Murphy-Hill, E. (2012, September 1). Something Smells Phishy: Exploring Definitions, Consequences, and Reactions to

Phishing. *Proceedings of the Human Factors and Ergonom-Ics Society Annual Meeting*, *56*(1). https://doi.org/10.1177/1071181312561447

Linkedin Cybertrust Latam. (2024). *La Nueva Era del Phishing: Cómo la Inteligencia Artificial Está Revolucionando los Ataques Cibernéticos*. https://www.linkedin.com/pulse/la-nueva-era-del-phishing-c%C3%B3mo-inteligencia-artificial-est%C3%A1-fdeme/

Lei, W., Hu, S., & Hsu, C. (2023, 06). Unveiling the process of phishing precautions taking: The moderating role of optimism bias. *Computers & Security*, *129*(103249). https://doi.org/10.1016/j.cose.2023.103249

Lei, W., Hu, S., & Hsu, C. (2023, 07 04). Uncovering the role of optimism bias in social media phishing: an empirical study on TikTok. *Behaviour & Information Technology*. ttps://doi.org/10.1080/0144929X.2023.2230305

Ley, W., Hu, S., & Hsu, C. (2023, 06). Unveiling the process of phishing precautions taking: The moderating role of optimism bias. *Computers & Security*, *129*(103249), 1 - 12. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4147323

Luo, X.R., Zhang, W., Burd, S., & Seazzu, A. (2013). Investigando la victimización por Phishing con el modelo heurístico-sistemático: un marco teórico y una exploración. *Computadoras y seguridad, 38,* 28-38. https://doi.org/10.1016/j.cose.2012.12.003

Medel, T., Schuster, R., Tromer, E., & Toch, E. (2022, 03). *Toward Proactive Support for Older Adults: Predicting the Right Moment for Providing Mobile Safety Help*. ACM Digital Library.

Michie, S., Atkins, L., & West, R. (2014). *The Behaviour Change Wheel*. Silverback Publishing.

Ogbanufe, O., & Pavur, R. (2023, 02). Going through the emotions of regret and fear: Revisiting protection motivation for identity theft protection. *International Journal of Information Management*, *62*(102432). https://doi.org/10.1016/j.cose.2023.103249

Pattinson, M., Butavicius, M., Parsons, K., McCormac, A., & Calic, D. (2015). Factors that Influence Information Security Behavior: An Australian Web-Based Study. *International conference on human aspectsof information security, privacy, and trust*, *Volume 9190*, (pp. 231–241). https://doi.org/10.1007/978-3-319-20376-8_21

Ramírez, M., Ruiz, S., & Osorio, M. C. (2022). *¿Qué tan fácil puedes caer en noticias falsas? Una intervención comportamental digital tipo boost para el fortalecimiento de capacidades de discernimiento de la información en escenarios electorales* [Universidad EAFIT]. https://repository.eafit.edu.co/items/0211a664-d37a-402e-8162-b1e1638fad50

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *Journal of Psychology: Interdisciplinary and Applied, 91* (1), 93-114.

Singh, K., Aggarwal, P., Rajivan, P., & Gonzalez, C. (2023, 02 7). Cognitive elements of learning and discriminability in anti-phishing training. *Computers & Security*, *127*(103105). https://doi.org/10.1016/j.cose.2023.103105

Sykosch, A., Wübbeling, M., & Doll, C. (2022, 08 25). *Analyzing User Behavior in Response to Controlled Stimuli for IT Security Awareness Assessment*. ACM Digital Library. org.ezproxy.eafit.edu.co/doi/pdf/10.1145/3407023.3409205

Tally, A. C., Abbott, J., Bochner, A., Das, S., & Nippert, C. (2023, 04). *What Mid-Career Professionals Think, Know, and Feel About Phishing: Opportunities for University IT Departments to Beter Empower Employees in Their Anti-Phishing Decisions*. ACM Digital Library. Retrieved 10 14, 2023, from https://dl.acm.org/doi/10.1145/3579547

Tarlogic. (2023, November 30). *Fraudes con Inteligencia Artificial: Nueva tecnología, viejos objetivos*. https://www.tarlogic.com/es/blog/fraudes-con-inteligencia-artificial/

Thomas R, P. (2006, Nov). Social Engineering: Concepts and Solutions. *Information Systems Security*, *Tomo 15*(5), 13-21. https://www.proquest.com/openview/6535856a33b27389b0f070f8a841c1bd/1?pq-origsite=gscholar&cbl=52433

TransUnion. (2023). *Informe 2023 sobre el fraude omnicanal*. https://www.transunion.co/content/dam/transunion/co/business/collateral/report/Reporte_Anual_Tendencias_Fraude_2023_Colombia.pdf

Valora Analitik. (2020, 08 6). *Nequi logró un crecimiento de 111 % en nuevos usuarios durante la pandemia*. https://www.valoraanalitik.com/2020/08/06/nequi-logr-un-crecimiento-de-111-en-nuevos-usu arios-durante-la-pandemia/

Vidas, T., Owusu, E., Wang, S., Zeng, C., Cranor, L., & Christin, N. (2013). QRishing: The Susceptibility of Smartphone Users to QR Code Phishing Attacks. *International Conference on Financial Cryptographyand Data Security*, *7862*, 52–69. https://doi.org/10.1007/978-3-642-41320-9_4

Wall, J. D., & Buche, M. W. (2017, 09 5). To Fear or Not to Fear? A Critical Review and Analysis of Fear Appeals in the Information Security Context. *AIS Journals*, *41*. 10.17705/1CAIS.04113

Weaver, B. W., & Braly, A. (2021). Training Users to Identify Phishing Emails. *Journal of Educational Computing Research*, *59*. 10.1177/0735633121992516

Williams, R., Morrison, B. W., Wiggins, M. W., & Smith, P. B. (2023, 07 05). *The role of conscientiousness and cue utilisation in the detection of phishing emails in controlled and naturalistic settings*. *Behaviour & Information Technology*. Taylor and Francis Online homepage.

Zeelenberg, M., & Pieters, R. (2007). A Theory of Regret Regulation 1.0. *Journal of Consumer Psychology, 17*, 29-35.

Zhuo, S., Biddle, R., & Kho, Y. S. (2024, 04). *Human-centered Phishing Susceptibility*. ACM Digital Library.

Anexos

Anexo 1. Instrumento de recolección de información

Tabla 11

Instrumento de recolección de datos de la población inicial

Guión de entrevista		
Texto introductorio		Me encuentro realizando una maestría en estudios del comportamiento y con fines académicos para la tesis de grado necesito hacer una investigación sobre seguridad online. Esta encuesta es completamente anónima y no tiene relación directa con la empresa. Todas las respuestas son válidas, te pido si decides ayudarme a responder con honestidad.
Tipo de pregunta	#	Pregunta
Demográficas	1	¿Cuál es tu identidad de género? Femenino Masculino Prefiere no decir
	2	Edad: Menor a 18 18 años a 24 años 25 años a 34 años 35 años a 44 años 45 años a 54 años Más de 54
	3	Estrato 1,2,3,4,5,6
	4	Lugar donde vive
	5	Nivel escolaridad: Primaria Bachiller Técnico Tecnólogo Profesional Posgrado
	6	Gerencia o área a la que perteneces

	7	¿A qué edad tuvo su primer dispositivo conectado a internet (Smartphone, Computador)? Menor a 18 18 años a 24 años 25 años a 34 años 35 años a 44 años 45 años a 54 años Más de 54
Antecedentes	8	¿Ha recibido información o capacitación sobre seguridad informática en el pasado? en caso que si, ¿dónde y cuál es el tema que más recuerdas?
	9	¿Usted o alguien en su círculo cercano (familia, amigos, colegas) ha sido víctima de algún robo, estafa de dinero o información por un medio digital en los últimos 12 meses (mensaje de texto, correo electrónico, redes sociales u otro? Si es así, ¿por qué cree que esto pasó?
	10	En una escala del 1 al 10, ¿qué tan preocupado está por su seguridad en línea, (Donde 1 es "no preocupado en absoluto" y 10 es "muy preocupado") ¿Por qué te sientes así?
Sesgo de optimismo	11	En una escala del 1 al 10, ¿Qué tan probable consideras que podrías llegar a ser víctima de robo de información o dinero por medios digitales en los próximos 12 meses? (Donde 1 es "no es nada probable" y 10 es "es muy probable") ¿Por qué?
	12	¿Has compartido tu información personal por links que te envían de manera electrónica y por qué?
Conocimiento	13	¿Sabes qué es phishing? Sí - No
	14	Entendiendo que el phishing es un ataque que consiste en engañar a las personas para robarle información confidencial o dinero por medios digitales ¿Como harias para identificarlo?
	15	¿Sabrías cómo y en dónde denunciar un intento de phishing?
	16	¿Qué factores consideras al evaluar la seguridad de un enlace o correo electrónico antes de interactuar con él?
Comportamiento de riesgo	17	En una escala de 1 a 10 ¿Qué tan probable es que des click a enlaces desconocidos?
	18	En una escala de 1 a 10 ¿Qué tan probable es que entregues información personal en línea? (por ejemplo, información de donde vives, de tus intereses, etc.)
	19	¿Hay situaciones en las que serías más propenso a hacerlo?

Nota. Elaboración propia, para analizar la población objetivo.

Anexo 2. Guías de Michie

Tabla 12

Guías de Michie 1, problema

Descripción de lo que se necesita identificar	Pregunta	Respuesta
El comportamiento o los comportamientos (lo más específico posible) que necesitan ser cambiados para solucionar un problema.	¿Qué comportamiento? La redacción inicia con un verbo en infinitivo	Identificar que el mensaje contiene alertas de phishing, y por ende no ejecutar la acción sugerida.
Las características del lugar en el que ocurre el comportamiento.	¿Dónde ocurre el comportamiento?	La conducta puede ocurrir en cualquier lugar físico y en cualquier ambiente digital, mediante un dispositivo electrónico que pueda recibir mensajes en alguna plataforma como las mencionadas. De hecho, esta característica dificulta las intervenciones pues no hay un lugar centralizado donde suceda el comportamiento.
Los individuos, grupos o poblaciones que están involucradas en el comportamiento.	¿Quiénes están involucrados en la realización del comportamiento?	Existen dos actores principales y uno secundario: En primer lugar está la víctima del ciberdelito, que recibe el mensaje y debe decidir entre entrar al link y proporcionar información o no hacerlo. Está también el atacante que diseña el delito y está atento a las víctimas que caigan en él. Existe además un tercer actor que es la institución asociada (banco, red social, etc) en cuyo nombre se diseña el ataque. Este actor si bien es pasivo en el ataque en sí, puede sufrir de pérdida reputacional y adicionalmente es un actor fundamental en el reporte y atención a la víctima del ataque. Existen otros actores involucrados encargados de habilitar la tecnología que permite el ataque, como lo son las empresas de telecomunicaciones (proveedores de mensajes de texto) o plataformas de correo electrónico (Gmail, Outlook, etc.), y las empresas que proveen los sistemas operativos de los celulares (Apple, Google, etc.)

Nota. Elaboración propia, adaptado de *The Behaviour Change Wheel,* por S. Michie et al., 2014, Silverback Publishing.

Tabla 13

Guías de Michie análisis del foco comportamental

Propósito de la intervención: Hacer que las personas logren identificar señales de phishing en mensajes de texto y en correo electrónico

Lista de comportamientos potenciales

a. La víctima se educa en los riesgos y características del P*hishing*
b. La víctima educa a otras personas a su alrededor sobre estos riesgos.
c. La institución asociada (banco, red social, etc.) educa a sus usuarios sobre los riesgos del *Phishing*, y también en cómo identificar un mensaje seguro y uno que no lo es.
d. La víctima abre el mensaje pero identifica rasgos que le generan desconfianza (URL, número de teléfono del remitente, entre otros), y no abre el link.
e. La víctima elimina el mensaje

Comportamientos potenciales	Impacto	Probabilidad de cambio	Efecto derrame	Facilidad de medición
a. La víctima se educa en los riesgos y características del *Phishing*	Prometedora	Prometedora	Muy prometedora	Poco Prometedora pero considerable
b. La víctima educa a otras personas a su alrededor sobre estos riesgos.	Prometedora	Poco Prometedora pero considerable	Poco Prometedora pero considerable	Inaceptable
c. La institución asociada (banco, red social, etc.) educa a sus usuarios sobre los riesgos del Phishing, y también en cómo identificar un mensaje seguro y uno que no lo es.	Prometedora	Prometedora	Poco Prometedora pero considerable	Muy Prometedora
d. La víctima abre el mensaje pero identifica rasgos que le generan desconfianza (link, número de teléfono del remitente, entre otros), y no abre el link.	Prometedora	Prometedora	Muy prometedora	Prometedora
e. La víctima elimina el mensaje	Muy prometedora	Muy prometedora	Muy prometedora	Inaceptable
Conducta objetivo seleccionada	La víctima abre el mensaje pero identifica rasgos que le generan desconfianza (link, número de teléfono del remitente, entre otros), y no abre el link.			

Nota. Elaboración propia, adaptado de *The Behaviour Change Wheel,* por S. Michie et al., 2014, Silverback Publishing.

Tabla 14

Guías de Michie análisis del foco comportamental 2

Foco de comportamiento: La víctima abre el mensaje pero identifica alertas de phishing y no ejecuta la acción	
¿Quién necesita llevar a cabo el comportamiento?	Potencial víctima
¿Qué acciones diferentes debe realizar para lograr el cambio deseado?	- Mantenerse informado en las diferentes técnicas de Phishing y cómo identificarlo. - Al recibir un mensaje, la víctima debe analizarse para confirmar si se encuentra en un momento adecuado para detectar un mensaje fraudulento. - Al recibir un mensaje, evaluar elementos de su contenido para verificar la veracidad de la fuente antes de ejecutar alguna acción. - En caso de tener dudas, comunicarse directamente con la entidad para validar el mensaje. - no ejecutar la acción sugerida por el mensaje - Eliminar el mensaje
¿Cuándo debe hacerlo?	Cuando reciba el mensaje
¿Dónde debe hacerlo?	En el dispositivo electrónico que llegue el mensaje
¿Con qué frecuencia debe hacerlo?	Cada que reciba un mensaje
¿Con quién(es) más deben llevar a cabo el comportamiento?	Con nadie

Nota. Elaboración propia, adaptado de *The Behaviour Change Wheel,* por S. Michie et al., 2014, Silverback Publishing.

Tabla 15

Componentes del COM-B y su necesidad de cambio

Componentes del COM-B	¿Qué necesita cambiar para que la conducta objetivo ocurra?	¿Hay necesidad de cambio?
Capacidad física	Acceso a un dispositivo con capacidad para recibir y enviar mensajes, y para acceder a internet de forma segura.	No
	Manejo adecuado de dispositivos móviles, ya que son aproximadamente tres veces más vulnerables a los ataques de phishing debido a varias características de los dispositivos móviles, incluido el pequeño tamaño de la pantalla, la inconveniencia de la entrada del usuario y la falta de indicadores de seguridad (Goel & Jain, 2018; Vishwanath, 2016).	No
Capacidad psicológica	La habilidad para manejar dispositivos electrónicos y navegar por internet	No
	Conocimiento de lo que es phishing, cómo identificarlo y el riesgo que conlleva (Greitzer et al., 2021)	Si

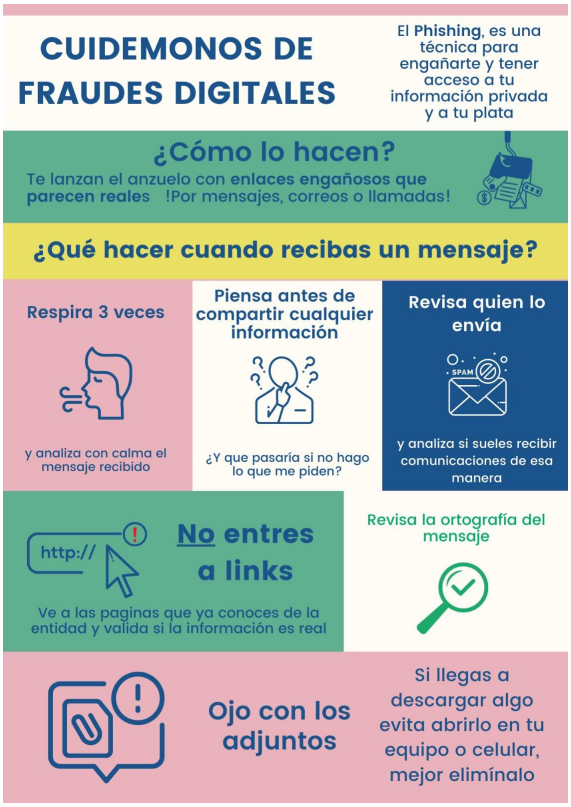
	Habilidades de análisis crítico y/o atajos para evaluar la legitimidad de los mensajes de manera efectiva. (Williams et al., 2023)	Si
	Esfuerzo cognitivo: la cantidad de esfuerzo cognitivo que dedican a comprender el mensaje contribuye directamente a su rendimiento. En este contexto, el esfuerzo cognitivo incluye la conciencia, la atención y la elaboración. (Greitzer et al., 2021)	Si
	Capacidad de asumir una postura de prevención de riesgos adecuada al abordar mensajes electrónicos. La predisposición al riesgo de las personas puede influir en cómo actúan al recibir un mensaje de phishing (Abroshan et al., 2021).	Si
Oportunidad física	Disponibilidad de recursos en línea para verificar la legitimidad de las instituciones (por ejemplo, líneas telefónicas oficiales, páginas web oficiales). En numerosos portales que consultamos la recomendación más común ante sospecha de phishing es acceder a un canal oficial de la institución (Grupo Bancolombia, 2017).	Si
Oportunidad social	No hemos encontrado literatura ni intervenciones que sustenten una oportunidad de cambio	No
Motivación reflexiva	Generar autoconciencia de estar expuesto a un ataque de phishing. El sesgo de optimismo se correlaciona con una mayor susceptibilidad de ser víctima de phishing (Ley et al., 2023). Por lo tanto, generar una reacción automática de prevención ayudará a lograr la conducta objetivo.	Si
	Aversión al arrepentimiento. Se puede capitalizar la sensación de arrepentimiento al ser víctima de phishing, esto es una emoción negativa que los individuos experimentan cuando “nos damos cuenta o imaginamos que nuestra situación actual habría sido mejor si hubiéramos decidido de manera diferente” (Zeelenberg y Pieters, 2007, p. 3). Es una emoción que la mayoría de la gente intenta evitar o minimizar	
Motivación Automática	Motivación a responder rápidamente ante amenazas (Rogers, 1975). (Teoría de la motivación de protección PMT). Apelar al miedo es un mecanismo efectivo para generar más prevención a la hora de encarar medios electrónicos.	Si
	La satisfacción de sentirse seguro y protegido también.	Si

Nota. Elaboración propia, adaptado de *The Behaviour Change Wheel,* por S. Michie et al., 2014, Silverback Publishing.

Anexo 3. Infografía de la intervención

Figura 7

Infografía entregada en la intervención

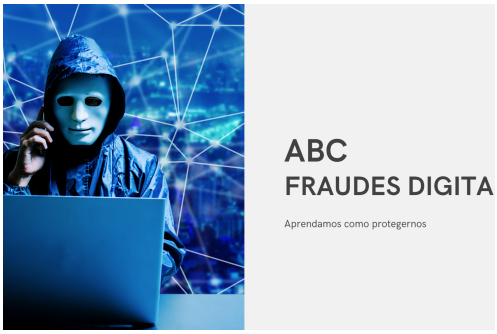


*Nota***.** Elaboración propia para la intervención

Anexo 4. ABC del phishing

Figura 8

Presentación utilizada dentro de la intervención



ABC
FRAUDES DIGITA

Aprendamos como protegernos

¿Por qué es un problema y es importante aprender a protegernos?

Durante el 2022, las denuncias en Colombia por ciberdelitos cre un 26%

Cada 8 minutos se registró una nueva denuncia en el país

El hurto por medios informativos es el delito mas frecuente: un 25.413

Los estafadores cada vez más logran hacer mensajes difíciles de detectar volviendonos más vulnerables

Puedes perder información personal, dinero afectando tu reputa tus finanzas

El Phishing es como salir de pesca. Donde tu eres e pez

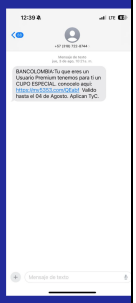
Los estafadores envían mensajes que parecen legítim (carnada) como un correo electrónico de tu banco, u social, o una tienda en línea, intentando que "muerd anzuelo". Te pueden pedir que hagas clic en un enla actualices tus datos, o incluso que confirmes tu cont

Pero una vez que proporcionas esa información, cae trampa, y los estafadores pueden usar tus datos par tu identidad, tu dinero o ambos.

Acciones urgentes

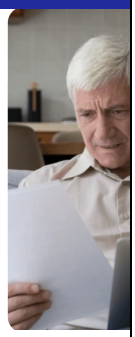
Te puede llegar un mail o un SMS solicitando realices una acción urgente.
Se hace pasar por una entidad conocida pidie que ingreses a un sitio falso aparentemente sim al sitio oficial de la entidad.

Acabas de recibir este mensaje, ¿tú creerías que es legítimo?



¿Qué es Phishing?

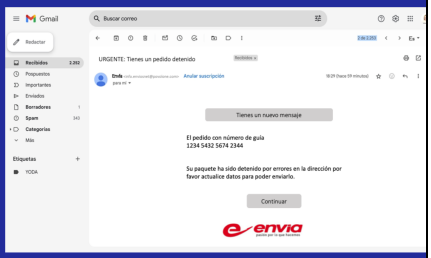
Imagínate que recibes una carta que parece ser de un banco pidiéndote que confirmes algunos datos personales. Pero, en realidad, la carta no es del banco, sino de alguien que quiere robar tu información. Esto, en el mundo digital, se llama "Phishing".



Argumentos que te dan los estafadores

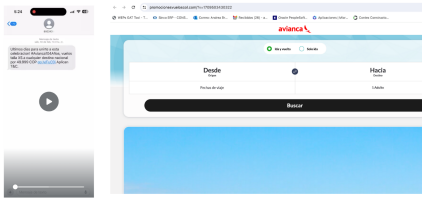
- Te piden acciones urgentes
- Te ofrecen beneficios exclusivos o promoci tentadoras
- Hacen que te de miedo, por posible perdid dinero, información o ambos.

Acciones urgentes



Promociones tentadoras

Te pueden ofrecer grandes descuentos, ofertas especiales, productos o servicios gratuitos bastante tentad



¿Cómo identificar si es phishing?

Revisa quién envía el mensaje o el correo Aunque pueda parecer legítima a primera vista, a menudo hay pequeñas diferencias o caracteres extraños que indican que es falso	Si tienen links, verifica realmente coincide con l oficial de la entid
Busca si el mensaje tiene errores gramaticales o de ortografía	Analiza el tono si es usualmente utiliza la e
Te da información irreal o desproporcionada (Si es demasiado bueno para ser real desconfía)	

¿Cuál de estas URLs te parece correcta?

https://bancolomb1a.com	https://bancolombia.com
https://netflix.com	https://netflix.misvideos.com
https://goo.gle.com	https://google.com
https://microsoft.com	https://micro.soft.com
https://microsoft.com/myprofile	https://microsoft.com/

Ejemplos reales

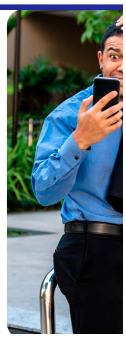
- El número del que es enviado, es de un proveedor d servicios de mensajes
- Tiene buena ortografía
- El link es corto y difícil de identificar su veracidad

Este mensaje no tiene una señal de Phishing por lo que difícil de identificar que es un engaño.

Qué acciones hacer?

- En caso de duda evita dar Clic en el enlace
- Si llegas a dar clic, cuando se abra la pagina valida e esta tiene, si es diferente a la del sitio oficial, NO ent información.
- Ve al sitio oficial o llama a la entidad y valida si la pr es real.

Juguemos



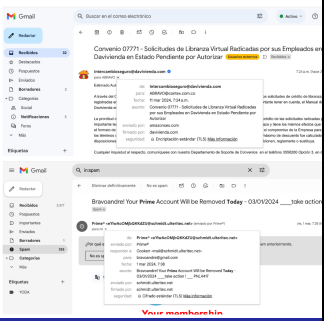
Generan miedo

El atacante te puede enviar un mensaje que te g miedo obligándote a realizar una acción rápidar y sin pensar.

Tips para identificar links (URL) sospechosas

✓	**Revisa la URL con atención:** los enlaces maliciosos tienen errores ortográficos o usan caracteres especiales enganñarte y hacer que parezcan legítimos. Por ejemplo, pueden usar "g00gle" "google".
✓	**Busca el candado de seguridad:** Al principio del URL, debe haber un icono de un candado que indica que la co segura. Si no está, podría ser un indicio de que el sitio no es seguro.
✓	**Verifica el dominio: (como termina la URL)** Desconfía de los sitios con terminaciones raras o que intentan imitar a otras p conocidas con cambios mínimos en el URL

Tips para identificar destinatarios de correos sospechosos

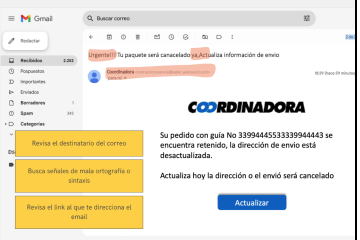


- El número del que se envía el mensaje un núm celular normal.
- Tiene mala ortografía
- El link no corresponde a la página oficial de A

Qué acciones hacer?

- Elimina el mensaje tiene todas las caracterís ataque de Phishing.

Este email tiene señales de alerta ante fraude?

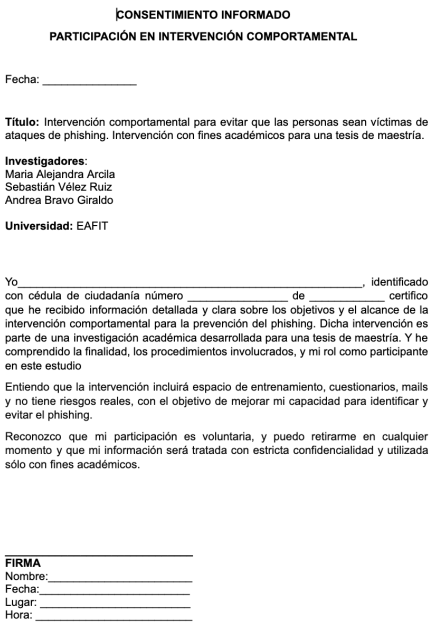


*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Anexo 5. Consentimiento Informado

Figura 9

Imagen del consentimiento informado utilizado para la intervención



CONSENTIMIENTO INFORMADO

PARTICIPACIÓN EN INTERVENCIÓN COMPORTAMENTAL

Fecha: _____

Título: Intervención comportamental para evitar que las personas sean víctimas de ataques de phishing. Intervención con fines académicos para una tesis de maestría.

Investigadores:
Maria Alejandra Arcila
Sebastián Vélez Ruiz
Andrea Bravo Giraldo

Universidad: EAFIT

Yo_____, identificado con cédula de ciudadanía número _____ de _____ certifico que he recibido información detallada y clara sobre los objetivos y el alcance de la intervención comportamental para la prevención del phishing. Dicha intervención es parte de una investigación académica desarrollada para una tesis de maestría. Y he comprendido la finalidad, los procedimientos involucrados, y mi rol como participante en este estudio

Entiendo que la intervención incluirá espacio de entrenamiento, cuestionarios, mails y no tiene riesgos reales, con el objetivo de mejorar mi capacidad para identificar y evitar el phishing.

Reconozco que mi participación es voluntaria, y puedo retirarme en cualquier momento y que mi información será tratada con estricta confidencialidad y utilizada sólo con fines académicos.

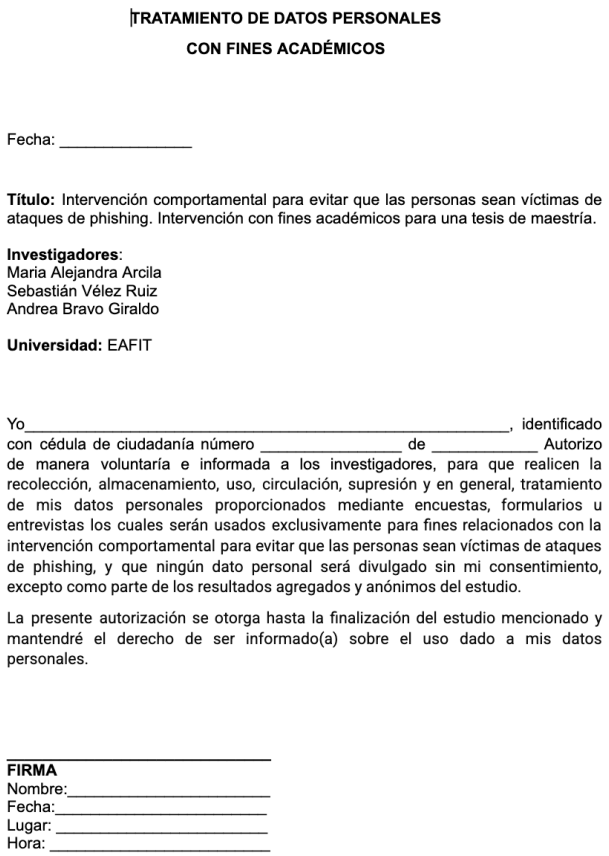
_____
FIRMA
Nombre:_____
Fecha:_____
Lugar: _____
Hora: _____

*Nota***.** Elaboración propia, con la encuesta demográfica realizada.

Anexo 6. Tratamiento de Datos Personales

Figura 10

Imagen del tratamiento de datos personales utilizado para la intervención



TRATAMIENTO DE DATOS PERSONALES
CON FINES ACADÉMICOS

Fecha: _____

Título: Intervención comportamental para evitar que las personas sean víctimas de ataques de phishing. Intervención con fines académicos para una tesis de maestría.

Investigadores:
Maria Alejandra Arcila
Sebastián Vélez Ruiz
Andrea Bravo Giraldo

Universidad: EAFIT

Yo_____, identificado con cédula de ciudadanía número _____ de _____ Autorizo de manera voluntaría e informada a los investigadores, para que realicen la recolección, almacenamiento, uso, circulación, supresión y en general, tratamiento de mis datos personales proporcionados mediante encuestas, formularios u entrevistas los cuales serán usados exclusivamente para fines relacionados con la intervención comportamental para evitar que las personas sean víctimas de ataques de phishing, y que ningún dato personal será divulgado sin mi consentimiento, excepto como parte de los resultados agregados y anónimos del estudio.

La presente autorización se otorga hasta la finalización del estudio mencionado y mantendré el derecho de ser informado(a) sobre el uso dado a mis datos personales.

_____
FIRMA
Nombre:_____
Fecha:_____
Lugar: _____
Hora: _____

*Nota***.** Elaboración propia para la intervención.