Servicio Nacional del Consumidor

FACULTAD ECONOMÍA Y NEGOCIOS UNIVERSIDAD DE CHILE

Fraude en Medios de Pago en Chile

Estudio Experimental:

¿Cómo reforzar las Campañas de prevención de las estafas en línea?

SERNAC
Subdirección de Consumo Financiero
Coordinación de Economía del Comportamiento

Universidad de Chile
Facultad de Economía y Negocios
Departamento de Administración

Santiago, julio de 2025

Equipos responsables de la publicación:

Equipo SERNAC:

Miguel Pavéz, Subdirector de Consumo Financiero (S);
Marcela Palominos, Coordinadora Economía del Comportamiento;
Guillermo Acuña, Analista Senior de la Coordinación de Economía del Comportamiento.

Equipo Facultad de Economía y Negocios de la Universidad de Chile:

Cristóbal Barra, profesor asistente de la Facultad de Economía y Negocios de la Universidad de Chile, Ph.D. in Marketing, University of South Carolina.
Ignacio Vargas, ayudante de Investigación Doctoral de la Facultad de Economía y Negocios de la Universidad de Chile, Ph.D.(c) in Marketing, University of North Texas.

SERNAC agradece la valiosa contribución en la realización de este estudio a Andrés Pavón (Ex Subdirector de Consumo Financiero, MSc Regulation, London School of Economics, MSc Public Policy, University College London); Carlos Noton (profesor asistente de la Escuela de Administración de la Universidad Católica de Chile, PhD in Economics, University of California, Berkeley); Daniel Schwartz (Profesor Asociado, Departamento de Ingeniería Industrial, Universidad de Chile, Ph.D. en Behavioral Decision Research, Carnegie Mellon University); Denise Laroze (Profesora Investigadora, Centro de Investigación en Complejidad Social, Universidad del Desarrollo, Ph.D. en Government, University of Essex).

Se agradece la asistencia de los siguientes estudiantes de la Facultad de Economía y Negocios de la Universidad de Chile, en calidad de pasantes: Dominga Atal, Ignacia Correa, Agustín Delgado, Natalia Gallo, Carlos García, Francisco Morales, Luis Ojeda, Pedro Olivares, Antonia Paris, Juan Parra, Vicente Perales, Diana Portugal, Ayelen Sandoval, Juan Pablo Sierralta y Gustavo Valladares.

Servicio Nacional del Consumidor

FACULTAD ECONOMÍA Y NEGOCIOS UNIVERSIDAD DE CHILE

Contenido

1. Introducción

En el entorno digital actual, **el fraude financiero cibernético se ha convertido en una amenaza significativa a nivel global**, afectando a individuos, empresas y gobiernos. Según el *World Economic Forum*, la ciberseguridad es uno de los riesgos globales más críticos a corto plazo (WEF, 2024). La digitalización de los servicios financieros ha acelerado tanto el progreso como la sofisticación de las técnicas utilizadas por los cibercriminales, quienes explotan vulnerabilidades en sistemas y usuarios mediante técnicas como *ransomware*, *phishing* e ingeniería social. Asimismo, la proliferación de dispositivos conectados con la Internet de las cosas (IoT) ha permitido que los delincuentes operen a una escala sin precedentes, aumentando la complejidad y la importancia de la ciberseguridad como un pilar fundamental en las estrategias de seguridad nacional e internacional (Edwards, 2024).

Las repercusiones de estas amenazas trascienden el ámbito digital, afectando tanto las finanzas como la confianza en las instituciones. Desde el robo de información personal hasta ataques contra infraestructuras críticas, las consecuencias incluyen pérdidas económicas sustanciales y graves impactos en la privacidad y seguridad personal. Además, la integración de tecnologías avanzadas como la inteligencia artificial y el aprendizaje automático ha incrementado la sofisticación de estas amenazas, haciéndolas más adaptables y difíciles de prever. Esto ha transformado el panorama de la seguridad, convirtiéndose en un desafío constante para individuos, organizaciones y gobiernos (Edwards, 2024).

Chile no ha sido ajeno a los crecientes desafíos en ciberseguridad, que afectan tanto a sus ciudadanos como a sus instituciones. Al respecto, las estadísticas sobre reclamaciones de fraudes financieros, presentadas por los consumidores a las entidades financieras, muestran un notable incremento en los últimos años, tanto en el número de casos como en los montos involucrados. Durante el segundo semestre de 2023, se alcanzó el nivel más alto de reclamaciones (457 mil), mientras que el primer semestre de 2024 registró el mayor monto total reclamado (US$235 millones). En el año 2023, los montos reclamados se duplicaron en comparación con el año anterior, alcanzando aproximadamente US$307 millones.

En este contexto, **la reciente Ley N° 21.673 publicada en el Diario Oficial el 30 de mayo de 2024, reformuló el marco regulatorio relacionado a la limitación de la responsabilidad de los usuarios de medios de pago con ocasión del fraude, contenido en la Ley N° 20.009**. Esta normativa, entre otros aspectos, estableció nuevas responsabilidades tanto para usuarios como para entidades reguladas. Por una parte, la reforma legal dispuso que "los usuarios deberán informarse y adoptar todas las medidas necesarias para prevenir el uso indebido, el fraude u otros riesgos afines a la utilización de los medios de pago a que se refiere esta ley y los mecanismos de autenticación asociados". Paralelamente, las entidades reguladas tienen el deber de proveer información periódica, clara, accesible y actualizada sobre las medidas de seguridad y las instrucciones para un uso seguro, fomentando prácticas responsables en la gestión de los medios de pago (art. 4 bis).

En este sentido, **la reforma destaca el papel fundamental del factor humano en la seguridad informática. Aunque la tecnología es esencial, por sí sola no es suficiente para prevenir los fraudes**, por lo que fomentar comportamientos a favor de la autoprotección de los consumidores, a través de diversas campañas de ciberseguridad, juega un rol fundamental (Furnell, et al.,2006; Liang & Xue, 2010; Ng Boon-Yuen, et al. 2009; Rhee, et al. 2009).

Sin embargo, no todas las entidades financieras ofrecen campañas específicas contra el fraude en medios de pago en línea en sus sitios web institucionales o a través de sus canales de comunicación con los clientes. En los casos en que sí las implementan, suelen ser parciales y no abarcan todos los tipos de fraudes a los que están expuestos los usuarios. Además, la información disponible suele ser incompleta y de difícil acceso (SERNAC, 2025a). Por otra

parte, no se han identificado evaluaciones que midan la efectividad de estas campañas en la prevención de fraudes, lo cual resulta problemático si consideramos que, según extractos de reclamos ingresados al SERNAC, los usuarios pueden llegar a interpretar de manera restrictiva y errónea las recomendaciones entregadas (SERNAC, 2025a).

En términos generales, las campañas de ciberseguridad tienen como objetivo principal sensibilizar a los usuarios sobre las amenazas cibernéticas, promover comportamientos seguros, proporcionar las habilidades y conocimientos necesarios a través de la capacitación y educación, así como influir en el cambio de actitudes y percepciones frente a la seguridad digital. Sin embargo, **a menudo las campañas suelen fracasar en promover comportamientos seguros de la población**, debido a una combinación de factores humanos y de diseño. Es por esta razón que las ciencias conductuales, aplicadas a las campañas de concientización, son esenciales para diseñar estrategias que promuevan, de manera efectiva, la adopción de hábitos precautorios para mitigar estos riesgos (Anderson et al., 2010; Acquisti et al., 2015).

A la vez, **dentro de los fraudes más recurrentes se encuentra el *Phishing***. El *phishing* es una forma de fraude que utiliza principalmente correos electrónicos, mensajes de texto o llamadas telefónicas para engañar a las personas. Mediante sofisticadas técnicas de ingeniería social y subterfugios técnicos, los atacantes buscan manipular a sus víctimas para que realicen acciones específicas, como revelar información financiera, credenciales de acceso o datos sensibles. La ingeniería social, por su parte, comprende un conjunto de estrategias diseñadas para influir en el comportamiento humano y lograr que las personas divulguen información confidencial o ejecuten acciones perjudiciales sin percatarse del engaño (APWG, 2024).

El presente estudio tiene como objetivo identificar y evaluar elementos comunicacionales que pueden optimizar la efectividad de las campañas contra el *phishing*. Se analiza cómo estos factores contribuyen, por un lado, a fortalecer la capacidad de las personas para clasificar correctamente los correos que recibe (entre fraudulentos y legítimos) y, por otro, a influir positivamente en sus actitudes y percepciones sobre la seguridad digital. Todo ello con el propósito de fomentar una mayor disposición a adoptar, de manera preventiva, las recomendaciones de seguridad.

Para este fin, se llevó a cabo un **Experimento Controlado Aleatorizado** (RCT, por sus siglas en inglés). Esta metodología, promovida por la OCDE como innovación en la gestión pública, destaca por su capacidad de generar evidencia confiable para predecir el impacto de políticas públicas y otras intervenciones (OECD, 2019). Bajo este marco metodológico, se realizaron evaluaciones de impacto tanto a nivel de tratamiento (***Cell Means Model***) como a nivel de factores (**ANOVA factorial**). Asimismo, se efectuaron pruebas de robustez, tanto paramétricas como no paramétricas, para confirmar la consistencia de los resultados. Finalmente, se llevaron a cabo análisis de heterogeneidad por género y grupo etario, con el propósito de determinar si los efectos de las intervenciones difieren entre distintos subgrupos.

Asimismo, el diseño experimental se fundamentó en **teorías de la comunicación y la persuasión**, como las propuestas por Hovland/Yale (1953) y McGuire (1969). Además, se incorporaron teorías que explican cómo el **cambio de actitudes influye en el comportamiento**, entre ellas la Teoría de la Acción Razonada, la Teoría del Comportamiento Planificado y el Modelo de Probabilidad de Elaboración Extendida, entre otras.

En concreto, se evaluó el impacto de tres estrategias comunicacionales que refuerzan una campaña estándar de detección de correos fraudulentos:

Mensajes motivacionales: Se trata de mensajes que refuerzan la autoeficacia y destacan la gravedad del *phishing*. Un estudio previo de SERNAC analizó qué factores motivan a los consumidores a protegerse contra el fraude informático, tomando como base la Teoría de

Motivación de Protección (TMP). Dicho estudio concluyó que la autoeficacia es el factor más determinante para fomentar la autoprotección, seguida de la percepción de la gravedad del fraude (SERNAC, 2025b).

Explicación de técnicas de ingeniería social: Ofrece una breve descripción de los métodos de manipulación empleados en correos fraudulentos, lo que facilitaría su identificación.

Enfoque de juego: Consiste en aplicar un test para evaluar la legitimidad de los correos, permitiendo a los participantes poner en práctica las recomendaciones brindadas y, a la vez, favorecer un proceso de persuasión comunicacional más efectivo, además de potenciar la inmersión en la narrativa de la campaña.

Estas estrategias comunicacionales buscan fomentar una mayor disposición de las personas a adoptar de forma preventiva las recomendaciones de seguridad. Para ello, se midieron las variables "**actitud hacia las medidas de detección de estafas en línea" e "intención de protegerse en el futuro**", ambas consideradas precursoras del comportamiento observado. Adicionalmente, se analizaron variables intermedias relacionadas con el proceso de persuasión (atención, comprensión, aceptación general hacia la campaña, autoeficacia y gravedad percibida), lo que permitió reforzar las conclusiones sobre la efectividad persuasiva de las intervenciones.

Los resultados a nivel de factores, muestran que el **"Juego" es el elemento más influyente para mejorar los indicadores de actitud e intención**, así como para las variables intermedias vinculadas al canal de persuasión. Su impacto es consistente y significativo desde el punto de vista estadístico, lo que subraya su potencial como herramienta clave para fortalecer la percepción y el comportamiento en contextos de ciberseguridad y protección contra fraudes.

Por su parte, la "**Explicación" también desempeña un rol importante al mejorar la aceptación general de la campaña y la percepción de autoeficacia**, aunque su efecto suele ser más notorio en ausencia del "Juego" o en combinación con la "Motivación".

Finalmente, **la "Motivación" por sí sola no mostró resultados significativos en la mayoría de las variables,** lo que indica que, aunque los mensajes motivacionales refuercen la percepción de gravedad y la autoeficacia, no son suficientes para producir cambios sustanciales en la percepción o el comportamiento preventivo. Sin embargo, cuando se combinan con la "Explicación", se observan mejoras significativas en su influencia, lo que sugiere que la efectividad de la "Motivación" depende de su adecuada integración con elementos informativos.

El documento se organiza de la siguiente manera: Tras la introducción del Capítulo 1, el Capítulo 2 aborda el marco teórico, donde se discuten los mecanismos de persuasión presentes en las campañas de concientización, así como la definición y los mecanismos de persuasión que operan a través de los juegos serios. En el Capítulo 3, se presenta la evidencia empírica sobre la falta de efectividad que muestran actualmente las campañas de ciberseguridad, y posteriormente se discuten los principales impactos de las intervenciones basadas en juegos en distintos procesos cognitivos de la persuasión. El Capítulo 4 describe en primer lugar la hipótesis causal del estudio, para luego detallar el diseño experimental utilizado. El Capítulo 5 expone los resultados principales, tanto del análisis a nivel de tratamiento como del análisis factorial, además de los resultados de las pruebas de robustez y heterogeneidad, junto con los tamaños de los impactos estimados. Finalmente, en el Capítulo 6 se exponen las conclusiones principales.

2. Marco Teórico

2.1 Mecanismos de Persuasión de las Campañas de Concientización

Las campañas de concientización, al igual que otras formas de comunicación, buscan ***influir o persuadir*** a las personas, con el fin de generar un cambio de comportamiento observable en ellas.

La **influencia** se entiende como cualquier modificación en los estados mentales, procesos psicológicos (cognitivos, afectivos, etc.) o comportamientos observables de las personas, provocada por un estímulo externo (Cialdini, 2009; Perloff, 2003; Pratkanis & Aronson, 1994). Dentro de este marco, la **persuasión** se distingue como un tipo específico de influencia orientado a cambiar la actitud hacia un estímulo determinado, ya sea un objeto, una persona o un tema (Petty & Cacioppo, 1986).

El **concepto de actitud** agrupa componentes afectivos, cognitivos y conductuales en una evaluación global sobre un estímulo. La persuasión tiene como objetivo modificar esta evaluación, ya sea alterando su aceptación (positiva o negativa) o ajustando su intensidad (de una postura polarizada a una más moderada, o viceversa). Cualquier cambio significativo en estas dimensiones se considera un cambio de actitud. En algunos casos, la transformación de una actitud, como constructo psicológico, puede generar modificaciones en el comportamiento observable, subrayando la conexión entre los cambios internos y sus manifestaciones externas (Petty er al, 2019).

Uno de los enfoques más influyentes en el estudio científico de la persuasión surgió en la Universidad de Yale a mediados del siglo XX, con Carl Hovland como figura principal, junto con colaboradores como Irving Janis y Harold Kelley (Hovland et al., 1953; Hovland & Janis, 1959). Basándose en teorías de la comunicación y el aprendizaje de la época, el **Modelo de comunicación y persuasión de Hovland/Yale** propuso que el cambio de actitudes ocurre mediante un proceso de aprendizaje que implica tres pasos: **atención** a la información relevante, **comprensión** del contenido y **aceptación** de las conclusiones. Además, el modelo destaca la influencia de factores como la credibilidad de la fuente, la estructura del mensaje y las características del receptor en la efectividad persuasiva. Para que el cambio de actitud se refleje en el comportamiento, en este modelo, el receptor debe prestar atención al mensaje, entenderlo y considerarlo válido para lograr un cambio de actitud, siendo esencial retener esta información en la memoria para que el cambio se refleje en el comportamiento.

Siguiendo esta línea, William McGuire (1969) desarrolló el **Modelo de Procesamiento de la información** que plantea la persuasión como el resultado de una secuencia de seis etapas: Exposición al mensaje, atención, comprensión, aceptación, retención y conducta. Este modelo, conocido como la "cadena causal de pasos de procesamiento", postula que un fallo en cualquiera de sus etapas disminuye la probabilidad de éxito persuasivo. Además, establece que la probabilidad de que un paso ocurra depende de la probabilidad acumulada de que se hayan cumplido todos los pasos previos, lo que se conoce como el principio de probabilidad conjunta. Posteriormente, McGuire reformuló su modelo en diversas ocasiones, ampliándolo hasta doce pasos o sintetizándolo en solo dos fases principales: recepción (atención y comprensión) y aceptación (McGuire, 1985).

Aunque existen diferencias notables entre los planteamientos de Hovland y colegas (1953) y el modelo de McGuire (1969), ambos coinciden en que el aprendizaje o la recepción del contenido de un mensaje constituye una condición previa necesaria para que se produzca la aceptación y, en consecuencia, el cambio de actitudes (Eagly & Chaiken, 1993). Sin embargo, esto no implica que el aprendizaje sea suficiente para garantizar la aceptación ni, mucho

menos, el cambio de actitudes. Es relevante señalar que, en ocasiones, se evalúa el éxito persuasivo de una campaña únicamente a partir de la atención y comprensión demostrada por los receptores, lo cual puede ser una medida limitada.

En otra línea, la investigación sobre el cambio de actitudes ha demostrado que la atención consciente y la comprensión de un mensaje persuasivo no siempre son necesarias para generar cambios en las actitudes. Mecanismos como asociaciones y procesos afectivos simples pueden influir en las actitudes incluso sin una comprensión completa del mensaje (Dijksterhuis, 2004; Zajonc, 1980). Esto sugiere que la comprensión, aunque central para la comunicación, no es un mediador indispensable ni suficiente para la persuasión (Fishbein y Ajzen, 1982; Greenwald, 1968). En otras palabras, aunque la comunicación y la persuasión a menudo coexisten, comprender un mensaje no garantiza un cambio de actitudes. Es posible compartir significados sin que ello implique convencer o persuadir.

Además, en el contexto de las campañas de concientización, incluso cuando se logra la aceptación y, por ende, la persuasión, no todo cambio de actitudes se traduce necesariamente en un cambio de conducta. Si bien las actitudes son un factor importante que influye en el comportamiento humano, existen otros elementos, tanto personales como situacionales, que también determinan las conductas (Ajzen & Fishbein, 2005; Fazio & Olson, 2014; Fazio & Zanna, 1981; Petty & Krosnick, 1995). Por ello, **para optimizar la predicción del comportamiento, resulta fundamental evaluar tanto las actitudes como otros factores relevantes, como la intención conductual.**

Asimismo, **al medir las actitudes, no basta con analizar únicamente su aceptación (favorabilidad) o intensidad; es igualmente importante considerar dimensiones como la ambivalencia y la fuerza de las actitudes**, ya que estas pueden mejorar significativamente la capacidad de predecir las conductas. La utilización de herramientas de medición adecuadas para evaluar las actitudes frente a una propuesta persuasiva resulta clave para estimar, de manera válida y fiable, si se ha logrado la persuasión (Blanco et al., 2017; Eagly & Chaiken, 1993; Petty, Fazio & Briñol, 2008).

Por otra parte, el **Modelo de Probabilidad de Elaboración** (ELM, por sus siglas en inglés), propuesto por Petty y Cacioppo (1986), sugiere que las personas buscan formarse actitudes válidas, lo que las motiva a procesar información para sentirse "en lo correcto". Sin embargo, **la motivación y la capacidad para procesar un mensaje persuasivo** varían según factores individuales y contextuales. El ELM identifica dos rutas principales hacia la persuasión:

Ruta central: Ocurre cuando el receptor tiene alta motivación y capacidad para reflexionar exhaustivamente sobre la información. En estas condiciones, el cambio de actitud se produce a través de argumentos sólidos y procesos psicológicos como: a) Evaluación de argumentos; b) Influencia en la dirección de los pensamientos generados (favorable o desfavorable), c) Validación de los pensamientos (metacognición). La metacognición es un avance fundamental del ELM, al centrarse en la evaluación de la certeza, relevancia y consistencia de los pensamientos generados durante este proceso persuasivo (validez cognitiva), así como en los aspectos emocionales asociados a estas ideas (validez afectiva) por parte del receptor (Petty et al., 2002, 2007).

Ruta periférica: Se activa cuando el receptor tiene baja motivación o capacidad para elaborar una reflexión exhaustiva. Aquí, el cambio de actitud depende de claves periféricas simples, como la credibilidad del emisor o asociaciones afectivas, sin análisis profundo de los argumentos. (Petty y Cacioppo, 1986; Petty et al., 2019; Petty y Wegener, 1998).
Finalmente, el ELM sugiere que, cuando la motivación y la capacidad del receptor lo sitúan en un nivel intermedio de elaboración (es decir, ni alta ni baja), diversas variables

relacionadas con el emisor, el mensaje o el contexto pueden influir en el grado de procesamiento de la información. Estas variables pueden aumentar o reducir la cantidad de pensamiento que el receptor dedica a analizar los argumentos del mensaje (Petty & Wegener, 1998).

En este contexto, las variables pueden afectar tanto la motivación como la capacidad del receptor para procesar deliberada y profundamente la información. Por ejemplo, factores como la relevancia personal del tema, la ambivalencia hacia el objeto de actitud o la **necesidad de cognición del receptor influyen en la motivación para procesar la información**. Por otro lado, elementos como el conocimiento previo sobre el tema, la complejidad del mensaje o la presencia de distracciones afectan la capacidad del receptor para procesar la información (Petty & Cacioppo, 1986). Este modelo ha sido fundamental para explicar los procesos psicológicos detrás de la persuasión, incorporando tanto factores cognitivos como metacognitivos y considerando las diversas circunstancias que afectan la eficacia de los mensajes persuasivos.

Por otra parte, existen diversos modelos que buscan explicar cómo las actitudes influyen en el comportamiento. Al respecto, los modelos más relevantes son la teoría de la acción razonada de Fishbein y Ajzen (1975), su ampliación conocida como la teoría del comportamiento planificado (Ajzen, 1991), el modelo compuesto propuesto por Eagly y Chaiken (1993, 1998) y el modelo MODE de Fazio (1990). Finalmente, se describe el modelo de la Teoría de Motivación de Protección (Rogers, 1975), que trata sobre las decisiones de protección de las personas en situaciones de riesgo y el Modelo de Proceso Paralelo Extendido (Witte, 1994).

La **Teoría de la Acción Razonada** fue desarrollada para predecir comportamientos razonados y deliberados, es decir, aquellos que son planificados de antemano. Según este modelo, el factor inmediato y determinante del comportamiento de los individuos es su intención, entendida como la motivación para actuar. En su conceptualización original, el modelo establece que las intenciones están influenciadas por dos factores principales: las actitudes y las normas subjetivas. El componente de actitud se refiere a la valoración personal del comportamiento, es decir, si la persona lo percibe como positivo o negativo. Por otro lado, las normas subjetivas representan la presión social percibida para realizar o abstenerse de realizar dicho comportamiento Fishbein y Ajzen (1975).

Posteriormente, Ajzen (1991) reconoció que las acciones también pueden están influenciadas por la percepción de las personas respecto a su propia capacidad para llevar a cabo una conducta determinada. Este concepto, conocido como **autoeficacia**, se define como la "convicción de que uno puede llevar a cabo con éxito la conducta requerida para lograr los resultados deseados" (Bandura, 1977). Al comprender cómo la autoeficacia puede impactar el comportamiento, Ajzen revisó la Teoría de la Acción Razonada para incluir el concepto de **control conductual percibido**. Este se refiere a las percepciones individuales sobre si se poseen los recursos y las oportunidades necesarias para realizar una determinada acción. La incorporación de este elemento condujo a la reformulación del modelo, que pasó a denominarse **Teoría del Comportamiento Planificado** (Ajzen, 1991; Ajzen y Madden, 1986).

La teoría de la acción razonada y la teoría del comportamiento planificado son los modelos de relaciones actitud-comportamiento que se han probado con más frecuencia. Las predicciones derivadas de los modelos han recibido un fuerte apoyo empírico. Básicamente, los resultados de los metaanálisis brindan evidencia sólida de que la teoría de la acción razonada y la teoría del comportamiento planificado son efectivas para predecir el comportamiento "reflexivo".

Sin embargo, las intenciones no siempre se convierten en acciones, y esto puede deberse a diversas razones. Entre ellas, la falta de motivación para ejecutar la conducta planificada, la indecisión sobre el momento adecuado para iniciarla, la carencia de conocimiento sobre cómo llevarla a cabo, o simplemente el olvido, a menudo provocado por distracciones (Sheeran y Webb, 2016).

Al respecto, cuando las intenciones se formulan y se miden como **"intención de implementación"** aumenta la probabilidad de que una persona actúe en consecuencia. (Gollwitzer, 1999, Gollwitzer y Brandstätter, 1997). En un nivel básico, las intenciones de implementación son como planes de "si-entonces", que especifican comportamientos que una persona necesitará realizar para alcanzar un objetivo (Sheeran, 2002). En otras palabras, las intenciones de implementación toman la forma de actitudes que hacen que un individuo se centre en especificar dónde, cuándo y cómo se llevará a cabo un comportamiento. Estas pueden presentarse en la forma de "Cuando me encuentre con la situación A, realizaré el comportamiento B".

El **Modelo Compuesto de relaciones actitud-comportamiento**, desarrollado por Alice Eagly y Shelly Chaiken (1993, 1998) establece un vínculo entre actitudes, intenciones y comportamiento. El Modelo Compuesto propone diversos factores que influyen en las actitudes hacia los comportamientos. Entre estos factores se encuentran: a) Hábitos, entendidos como respuestas automáticas basadas en comportamientos previos, b) Actitudes hacia objetivos, que reflejan la percepción sobre el propósito del comportamiento, c) Resultados utilitaristas, que se refieren a las recompensas o castigos asociados con la realización del comportamiento, d) Resultados normativos, relacionados con la aprobación o desaprobación social que podría generar la acción, e) Resultados de autoidentidad, que consideran cómo el comportamiento impacta el autoconcepto o la autoimagen del individuo. Eagly y Chaiken argumentan que algunos de estos factores pueden influir en las intenciones de actuar, mientras que otros pueden tener un impacto directo en el comportamiento. Este enfoque multidimensional permite capturar una visión más completa de las dinámicas actitud-comportamiento.

Finalmente, Russell Fazio (1990) desarrolló el **Modelo de Motivación y Oportunidad como Determinantes del comportamiento** (MODE, por sus siglas en inglés) para explicar las relaciones entre actitudes y comportamiento. Este modelo se clasifica como un modelo de proceso dual, ya que describe dos formas distintas en las que las actitudes pueden influir en el comportamiento. El modelo MODE plantea que, cuando los individuos tienen suficiente motivación y oportunidad, pueden basar su comportamiento en un análisis deliberado de sus actitudes y de la información disponible. Sin embargo, si la motivación o la oportunidad para tomar una decisión razonada es limitada, los individuos recurren a un procesamiento espontáneo de la información. En estas situaciones, la accesibilidad de la actitud desempeña un papel crucial, es decir, la facilidad con la que una actitud puede ser recuperada de la memoria. Si una actitud es altamente accesible, se activa automáticamente y guía el comportamiento de manera consistente con esa actitud. Por el contrario, cuando una actitud no es fácilmente accesible, no se activa automáticamente, lo que reduce su probabilidad de influir en el comportamiento.

Por otra parte, la **Teoría de la Motivación de Protección** (PMT, por sus siglas en inglés), desarrollada por Rogers (1975), explica los factores que influyen en las decisiones de protección de las personas en situaciones de riesgo (Rogers, 1975; Maddux & Rogers, 1983). La PMT propone que la respuesta a una amenaza depende de dos evaluaciones clave:
a) **Evaluación de la amenaza**: El individuo analiza la probabilidad y gravedad de la amenaza, incluyendo su vulnerabilidad personal. Una percepción alta de severidad y vulnerabilidad motiva a considerar acciones de protección. Si una persona considera que la amenaza es seria y se percibe como vulnerable a ella, es más probable que se sienta

motivada a tomar medidas preventivas. En primer lugar, debe ser consciente de la amenaza y evaluarla en consecuencia. A partir de esta evaluación, se inicia el proceso de afrontamiento, en el que la persona analiza posibles estrategias para reducir o mitigar el riesgo; b) **Evaluación de la respuesta**: Se evalúa la eficacia de las medidas recomendadas para mitigar la amenaza y la autoeficacia, es decir, la confianza en la propia capacidad para ejecutarlas. Estas percepciones determinan si una persona adoptará una acción específica de protección. Tras estas evaluaciones, los individuos deciden comportarse de forma adaptativa (protegiéndose de la amenaza) o no adaptativa (evitando acciones de protección).

El **Modelo de Proceso Paralelo Extendido** (EPPM, por sus siglas en inglés) analiza las consecuencias no deseadas de ciertas apelaciones, como los mensajes de miedo, cuando la amenaza percibida por un individuo supera su percepción de eficacia (eficacia de la respuesta y autoeficiencia). En un intento por manejar las emociones desagradables asociadas al miedo, las personas tienden a desacreditar el mensaje, interpretarlo como una manipulación o evitar reflexionar sobre la amenaza y las formas de prevenirla (Witte, 1992, 1994). El EPPM aborda el nivel de amenaza percibida, que se evalúa mediante dos componentes: la gravedad percibida (qué tan seria es la amenaza) y la susceptibilidad percibida (qué tan vulnerable se siente el individuo ante la amenaza). El EPPM predice que un miedo excesivo puede llevar a acciones contraproducentes, mientras que un equilibrio adecuado entre el miedo y la eficacia —que incluye la eficacia de la respuesta (creencia en que la solución propuesta es efectiva) y la autoeficacia (confianza en la capacidad personal para implementar la solución)— fomenta un comportamiento autoprotector.

2.2 Juegos serios como herramientas de persuasión

Enfoques basados en Juegos

Como se ha señalado previamente, las campañas de concientización tienen como objetivo influir en las actitudes y comportamientos de las personas para generar cambios positivos en la sociedad. En este contexto, los enfoques basados en juegos emergen como estrategias de persuasión innovadoras que pueden potenciar el impacto de las campañas, aumentando su efectividad en la promoción de comportamientos deseables.

Los enfoques basados en juegos se dividen principalmente en tres categorías: gamificación, juegos serios y aprendizaje basado en juegos, cada uno con propósitos y alcances específicos. La **gamificación** aplica elementos característicos de los juegos (puntos, niveles, recompensas y rankings) en contextos no lúdicos para motivar comportamientos y alcanzar objetivos. Su éxito se mide por el grado de disfrute del usuario, quien debe percibir la experiencia como un juego (Zichermann, 2010; Pelling, 2011; Deterding et al., 2011; Werbach & Hunter, 2013; Ripoll, 2014). Los **juegos serios** son herramientas interactivas diseñadas para reforzar conocimientos y desarrollar habilidades específicas en áreas como salud, marketing y ciberseguridad. Cada elemento del juego, desde la narrativa hasta las mecánicas, tiene un propósito educativo claro (Abt, 1974; Henderson et al., 2024). El **aprendizaje basado en juegos** (Game-Based Learning, GBL) utiliza juegos como medio de enseñanza, integrando su desarrollo en el proceso de aprendizaje. A través del juego, se facilita la adquisición de conocimientos o habilidades de manera activa y significativa, fomentando la participación y el compromiso de los estudiantes (Campbell y Kuncel, 2002; Wilson et al., 2009).
En el marco de las campañas de concientización, se centrará el análisis en el enfoque de los juegos serios.

Características de los juegos serios

Según Charsky (2010), los juegos serios incluyen mecanismos clave que facilitan el aprendizaje y la transformación de actitudes y comportamientos. A continuación, se describen los más destacados:

Competencia: Los juegos fomentan la competencia al activar el deseo de ganar entre los jugadores. Sin embargo, ganar por sí solo rara vez es una motivación suficiente. Para potenciar el comportamiento, los juegos integran objetivos y acciones directamente relacionados con el logro del jugador, promoviendo así un compromiso más significativo.

Metas: Los juegos estructuran experiencias en torno a objetivos específicos, proporcionando a los participantes información valiosa para resolver problemas futuros. Estos objetivos ofrecen a los diseñadores de juegos una oportunidad para promover momentos de reflexión profunda, permitiendo a los jugadores practicar habilidades y procesos vinculados con cambios particulares en actitudes o comportamientos. Además, la retroalimentación inmediata es fundamental para internalizar el aprendizaje (Gee, 2007), lo que a su vez puede generar niveles más profundos de participación.

Reglas: Las reglas en los juegos imponen restricciones sobre las acciones de los jugadores, estableciendo límites que rara vez pueden ser modificados. A un nivel superficial, estas reglas son esenciales para facilitar la sesión de juego al proporcionar instrucciones claras. Sin embargo, también desempeñan un papel crucial al representar realidades complejas (Alessi y Trollip, 2001). Al otorgar o limitar opciones, los juegos pueden simular situaciones de la vida real, como desigualdades estructurales o sistémicas (Wendorf Muhamad, 2019).

Contexto/Entorno: Más allá de estos mecanismos, los juegos serios también involucran a las personas en contextos que, de otra manera, podrían resultar inaccesibles (Carcioppolo et al., 2015). Por ejemplo, los entornos de juego permiten a los participantes ejecutar procesos cognitivos más avanzados y acceder a esquemas alternativos que normalmente no utilizan, todo ello en un espacio seguro que mitiga los riesgos asociados.

Para que los juegos serios faciliten procesos persuasivos, es esencial que cumplan con ciertos requisitos: (a) Crear experiencias estructuradas en torno a objetivos específicos que resulten útiles para la resolución de problemas futuros. (b) Fomentar la interpretación activa por parte de los participantes, de modo que el aprendizaje y las habilidades adquiridas se internalicen y sean aplicables en situaciones futuras. (c) Proporcionar retroalimentación inmediata durante la experiencia, lo que refuerza el proceso de aprendizaje y permite ajustes en tiempo real (Gee, 2007). A través de estas características formales, los juegos serios ofrecen un método interactivo para involucrarse con mensajes persuasivos que, de otra manera, podrían ser inaccesibles o difíciles de transmitir (Carcioppolo, et al, 2015).

Mecanismos de Persuasión de los juegos serios

Para comprender cómo los juegos serios pueden fomentar el proceso que conduce al cambio de comportamiento, se analizan los siguientes modelos: Teoría Cognitiva Social (SCT), Modelo de Probabilidad de Elaboración extendida (EELM), Modelo de Superación de Resistencia del Entretenimiento (EORM) y Teoría de la Autodeterminación (SDT).

En primer lugar, de acuerdo a la **Teoría Cognitiva Social** (SCT, por sus siglas en inglés) propuesta por Bandura (1986), observar a otros realizar un comportamiento específico puede motivar a los individuos a replicarlo, al influir directamente en su autoeficacia, es

decir, la creencia en su capacidad para ejecutar una tarea o alcanzar un objetivo. Este mecanismo desempeña un papel esencial en la promoción de cambios conductuales.

El **Modelo de Probabilidad de Elaboración Extendida** (EELM, por sus siglas en inglés), propuesto por Slater y Rouner en 2002, es un enfoque diseñado para comprender cuándo y por qué se imitan los comportamientos observables. Basado en la teoría cognitiva social, este modelo postula que el aprendizaje se refuerza a través de la observación. Sin embargo, no todas las conductas observadas se adoptan. La motivación —o la falta de ella— desempeña un papel fundamental, ya que la exposición por sí sola no garantiza la adopción del comportamiento observado.

El EELM analiza los factores que podrían influir en la motivación y su impacto en la adopción de actitudes o conductas observadas. El modelo investiga, por ejemplo, cómo la resistencia hacia los mensajes puede obstaculizar la motivación para adoptar el comportamiento observado (Slater y Rouner, 2002; Moyer-Gusé, 2008). Según el EELM, cuando se presenta una narrativa persuasiva a través de un juego, aumenta la implicación del individuo (quien se ve inmerso en la historia y se torna menos crítico con el contenido), lo que a su vez reduce los contraargumentos y la resistencia al mensaje (Shrum, 2004; Slater y Rouner, 2002). Además, EELM postula que ciertos constructos, como la identificación, la homofilia y la transportación, permiten que este proceso ocurra.

a) Identificación

Un concepto relacionado con la adopción de roles es la identificación con personajes, que implica la colocación intencional y temporal del individuo en la posición de un personaje clave para fomentar una mayor comprensión (Flavell, et al., 1968; Kelley el al., 1975). **La identificación se caracteriza por la inmersión del jugador en el entorno del personaje, permitiendo una comprensión tanto cognitiva como emocional**. Este proceso facilita respuestas emotivas, como la empatía, que pueden ser transformadoras y motivar a los individuos a alinear sus objetivos con los del personaje (Cohen, 2001). La identificación involucra un compromiso mental desde una perspectiva de ideología suspendida (Tal-Or y Cohen, 2010) y se compone de cuatro dimensiones únicas: (a) Sentimiento compartido (empatía): experimentar las emociones del personaje; (b) Cogniciones compartidas: alinear el pensamiento cognitivo con el del personaje; (c) Objetivos compartidos: motivarse para actuar como el personaje; (d) Absorción: transportarse a la narrativa de la historia. Estas dimensiones permiten una conexión más profunda con el personaje, alentando a los individuos a responder desde la perspectiva del rol asignado o elegido, en lugar de actuar según sus respuestas habituales (Cohen, 2001). Este proceso de distanciamiento de los esquemas preexistentes posibilita una comprensión más matizada de las experiencias ajenas. Además, esta comprensión reflexiva contribuye a reducir las barreras de resistencia, como el miedo, colocando a los individuos en una posición de mayor apertura hacia actitudes y comportamientos prosociales integrados en el contenido narrativo persuasivo (Slater y Rouner, 2002). Como advertencia, Tal-Or y Cohen (2010) encontraron que los miembros de la audiencia mostraban una menor aceptación hacia personajes altamente estigmatizados. En contraste, interpretar personajes con estigmas moderados generó reacciones menos negativas. Esto llevó a los autores a concluir que la adopción de roles puede verse limitada cuando el nivel de estigmatización es excesivamente alto, lo que dificulta el proceso de identificación.

b) **Homofilia**

La homofilia, o similitud percibida, se refiere a la **creencia de un individuo de que comparte conocimientos, emociones y/o metas con un personaje clave.** En esencia,

la persona percibe que podría formar parte de la historia contextual como uno de los personajes principales. A diferencia de la identificación, la homofilia no implica una experiencia indirecta, como vivir las emociones de otro, sino que se basa en la conexión a través de atributos compartidos, como preferencias o aversiones similares.

c) Transporte

Mientras que la identificación y la homofilia se centran en la interacción del individuo con un personaje, **el transporte se enfoca en la interacción con el contenido narrativo**. Este concepto se define comúnmente como la absorción completa de un individuo, en la cual la persona logra "perderse" en la historia (Green y Brock, 2000). Durante este estado, el enfoque ya no recae en la elaboración o el procesamiento consciente del contenido narrativo, sino en una inmersión total en la experiencia (Gerrig, 1993).

En el contexto de los juegos serios, el transporte facilita momentos de "suspensión de la incredulidad" (Gilbert, 1991), en los cuales las historias, ficticias o no, cobran vida para los participantes. Durante estos momentos, la resistencia al mensaje disminuye y los individuos se vuelven más receptivos a los intentos de persuasión. Esto ocurre porque el contraargumento, una forma de resistencia activa, es incompatible con el estado de transporte: no es posible estar completamente absorbido en la narrativa mientras se generan contraargumentos (Moyer-Gusé, 2008). En los juegos serios, la absorción no solo mejora los esfuerzos cognitivos para comprender y experimentar la perspectiva del otro (es decir, un personaje clave), sino que también contribuye a reducir barreras como la resistencia al mensaje, los contraargumentos y la percepción de invulnerabilidad. Esto hace del transporte una herramienta clave para fomentar la conexión emocional y cognitiva con el contenido narrativo.

En este sentido, la identificación con los personajes en los juegos serios permite al jugador experimentar emociones y situaciones de manera más directa, lo cual motiva cambios de actitud. Esta identificación, a su vez, se ve reforzada por la homofilia y el transporte, que permiten a los jugadores sentir que comparten objetivos con los personajes y/o se sumergen completamente en la narrativa del juego.

Por otra parte, aunque el modelo de probabilidad de elaboración extendida (EELM) reconoce el papel de la resistencia psicológica como un factor que limita los esfuerzos persuasivos, no aborda las causas subyacentes de dicha resistencia. Para llenar este vacío, **el Modelo de Superación de Resistencia del Entretenimiento** (EORM) proporciona un marco para comprender las causas de la resistencia al mensaje y su efecto en la adopción de actitudes y conductas prosociales (Moyer-Gusé, 2008). La resistencia al mensaje se define como un fenómeno psicológico en el que los individuos rechazan mensajes persuasivos debido a una amenaza percibida o a otros factores, como la franqueza, el tono autoritario o el carácter intrusivo del mensaje (Buller, Borland y Burgoon, 1998; Knowles y Linn, 2004). Según Brehm (1966), esta resistencia surge de la necesidad del individuo de preservar su sensación de libertad en la toma de decisiones. Cuando un mensaje se percibe como restrictivo o sofocante, puede generar un efecto bumerán conocido como reactancia, es decir, una respuesta opuesta a la intención persuasiva del mensaje.

El EORM plantea que los mensajes enmarcados dentro de contextos de entretenimiento tienen menos probabilidades de provocar resistencia (Moyer-Gusé, 2008). Además, sugiere que cuando la reactancia es baja, los mensajes prosociales logran un mayor impacto, fomentando cambios significativos en actitudes y comportamientos (Moyer-Gusé, 2008). Según este enfoque, la reactancia frente a una amenaza percibida puede mitigarse al aumentar el disfrute narrativo. La creación de entornos narrativos ricos y complejos, que minimicen la percepción de limitación de la libertad de elección, reduce la probabilidad de

generar resistencia psicológica o reactancia. De esta manera, dichos entornos permiten que los mensajes persuasivos sean más efectivos (Moyer-Gusé, 2008).

La **Teoría de la Autodeterminación** (SDT, por sus siglas en inglés), desarrollada por Deci y Ryan (1985, 2001), explica la motivación humana a partir de la satisfacción de tres necesidades psicológicas fundamentales: **autonomía** (sentirse en control de las propias acciones), **competencia** (percibirse capaz y eficaz) y **relación social** (sentirse conectado con los demás). Según la SDT, las personas poseen una motivación intrínseca natural para actuar de manera autónoma y efectiva en su entorno, aunque esta puede verse influenciada por factores sociales y ambientales.

La teoría distingue entre motivación intrínseca y extrínseca. La **motivación intrínseca** impulsa a realizar una actividad por el placer o interés que esta genera en sí misma, mientras que la motivación extrínseca responde a incentivos externos como recompensas o sanciones. Sin embargo, la SDT propone que la motivación extrínseca puede internalizarse, permitiendo que las personas adopten conductas de forma autónoma. Dentro de la motivación extrínseca, se identifican cuatro niveles según el grado de internalización: a) Regulación externa: la conducta se basa en recompensas o castigos externos; b) Regulación introyectada: la acción se motiva por culpa, obligación o necesidad de aprobación; c) Regulación identificada: la persona reconoce la actividad como valiosa y alineada con sus propios objetivos; d) Regulación integrada: los valores y metas de la actividad se internalizan completamente, guiando el comportamiento de manera autónoma. Finalmente, la motivation ocurre cuando no hay intención de actuar, generalmente debido a la falta de control o sentido en la tarea.

3. Marco Empírico

3.1 Eficacia de las Campañas de Ciberseguridad

En términos generales, las campañas de ciberseguridad tienen como objetivo principal sensibilizar a los usuarios sobre las amenazas cibernéticas, promover comportamientos seguros, proporcionar las habilidades y conocimientos necesarios a través de la capacitación y educación, así como influir en el cambio de actitudes y percepciones frente a la seguridad digital. Sin embargo, **a menudo estas campañas suelen fracasar en promover comportamientos seguros de la población, debido a una combinación de factores humanos y de diseño**.

Desde la perspectiva humana, muchos usuarios carecen de los conocimientos y habilidades necesarias para protegerse eficazmente frente a amenazas digitales (Hong & Furnell, 2021; Tan & Aguilar, 2012). Además, la creación de hábitos inseguros, como el uso de contraseñas débiles o la omisión de actualizaciones de seguridad, incrementa significativamente la exposición a riesgos (Hong & Furnell, 2021). Otro aspecto crítico es la evaluación incorrecta de los riesgos: los usuarios tienden a subestimar la gravedad y probabilidad de los ataques, al tiempo que sobreestiman sus propias capacidades para enfrentarlos (Hong & Furnell, 2021; Tan & Aguilar, 2012; Frank et al., 2023). La percepción de que las medidas de seguridad requieren un esfuerzo excesivo también actúa como una barrera, desmotivando a los usuarios a implementarlas (Hong & Furnell, 2021; Grobler et al., 2021). El estrés y la presión temporal, comunes en el ámbito de la ciberseguridad, agravan la situación, ya que pueden llevar a tomar atajos inseguros o ignorar advertencias críticas (Brilingaitė et al., 2025). Además, la delegación de responsabilidades en terceros diluye la percepción de control personal sobre la seguridad (Tan & Aguilar, 2012; Hoiland, 2023). Incluso factores sociales, como el temor a ser percibido como paranoico o deshonesto al seguir prácticas de ciberseguridad, pueden disuadir a los usuarios de adoptar conductas seguras (Das, 2017).

Por otro lado, las **deficiencias en el diseño de las campañas** también pueden conducir a su fracaso. Las medidas de seguridad suelen percibirse como complicadas e incómodas, y con frecuencia se presentan de manera que sobrecargan al usuario con exceso de información, lo que termina desalentando su adopción (Hong & Furnell, 2021). Además, las campañas no siempre explican adecuadamente la sofisticación de técnicas como el *phishing*, dejando a los usuarios vulnerables a este tipo de manipulaciones (Hong & Furnell, 2021). Las asimetrías de información entre proveedores y usuarios también son un problema: los primeros no siempre comunican el nivel real de seguridad de sus productos, mientras que los usuarios confían en que las empresas mitigarán las vulnerabilidades existentes (Fonfría & Duch-Brown, 2020). Otra barrera significativa es la desconexión entre el diseño técnico de los sistemas y las necesidades de los usuarios. Las medidas de seguridad, como los cambios obligatorios de contraseñas o la configuración de autenticación multifactor, suelen percibirse como intrusivas, lo que fomenta atajos inseguros, como reutilizar contraseñas (Grobler et al., 2021). Además, muchas campañas carecen de métricas claras para evaluar su impacto y de un refuerzo continuo que permita cambios graduales en actitudes y comportamientos. Las campañas de corta duración o sin seguimiento pueden perder impacto con el tiempo (Bada et al., 2019). También se ha observado que el uso ineficaz del miedo como táctica puede ser contraproducente si las amenazas se perciben como irrelevantes o generan rechazo emocional (Bada et al., 2019). Además, cuando los mensajes son percibidos como impositivos o autoritarios, es probable que los destinatarios experimenten reactancia psicológica, resistiéndose activamente al cambio propuesto (Moyer-Gusé, 2008).

Para abordar estas limitaciones, se sugiere un enfoque que combine capacitación continua, diseño intuitivo y personalización cultural. La formación constante y la repetición de comportamientos deseados pueden convertirlos en hábitos automáticos, fortaleciendo la adherencia a las políticas de seguridad (Nord et al., 2020). Además, diseñar sistemas de ciberseguridad accesibles y comprensibles para usuarios con diferentes niveles de conocimiento técnico puede reducir la percepción de complejidad (Grobler et al., 2021). Por último, adaptar los mensajes de las campañas a contextos específicos aumenta su relevancia y efectividad, lo que contribuye a mitigar las barreras identificadas y a promover comportamientos seguros a largo plazo (Bada et al., 2019).

3.2 Principales impactos de las intervenciones basadas en juegos

Existe una amplia literatura empírica que demuestra que los métodos basados en juegos suelen ser más efectivos que los enfoques tradicionales para mejorar el aprendizaje, el desarrollo de habilidades, la transformación de percepciones y actitudes, y, en última instancia, la modificación de conductas en los participantes (Prümmer et al., 2025; Hammady et al., 2022; Egashira et al., 2022; Yildirim, 2017; Ezezika et al., 2018; Burkey et al., 2013). Esto se debe, en gran medida, a los efectos positivos que los enfoques lúdicos generan en distintos aspectos del proceso de persuasión: a) **Emocionales:** aumentan la satisfacción, el disfrute y el interés; reducen comportamientos impulsivos, la ansiedad y la percepción de dificultad. (Khan et al., 2023; Türkmen & Soybaş, 2019; Smith, 2017; Morillas et al., 2016); b) **Motivacionales**: promueven la motivación intrínseca, el compromiso, la autoeficacia y el control percibido. Además, mejoran la percepción de facilidad de uso y utilidad de los temas y fomentan actitudes más positivas hacia el aprendizaje (Khan et al., 2023; Hammady et al., 2022; Boncu et al., 2022; Yildirim, 2017; Muhamad & Kim, 2020; Burkey et al., 2013; Morillas et al., 2016; Buckley & Doyle, 2016); c) **Cognitivos**: potencian la atención, el pensamiento lógico y la reflexión crítica; favorecen una mejor evaluación del riesgo y la relevancia de los temas; mejoran la retención del conocimiento y la capacidad para identificar y corregir errores conceptuales. (Mota et al., 2016; Boncu et al., 2022; Muhamad & Kim, 2020; Morillas et al., 2016; Smiderle et al., 2020; Batzos et al., 2023; Bitrián et al., 2024); d) **Relacionales**: fortalecen las interacciones y la participación entre los usuarios, además de incentivar la promoción de buenas prácticas entre sus círculos cercanos. (Mota et al., 2016; Türkmen & Soybaş, 2019; Ezezika et al., 2018; Burkey et al., 2013).

En el ámbito de la ciberseguridad, con los métodos basados en juegos se observaron mejoras en la concientización y en las habilidades de los usuarios (Röpke, 2023; Bitrián et al., 2024). Los juegos facilitaron la visualización de las consecuencias inmediatas de las acciones de los participantes, lo que fomentó comportamientos más alineados con las políticas de seguridad (Alkhazi et al., 2022; Yasin et al., 2019). Además, incrementaron el disfrute, el interés en participar en futuras capacitaciones y la disposición para compartir conocimientos con compañeros. También contribuyeron a reducir la fatiga de seguridad y a promover una conciencia sostenida sobre los riesgos, así como una mayor atención, comprensión y motivación (Alkhazi et al., 2022; Yasin et al., 2019). Sin embargo, algunas experiencias no lograron los resultados esperados en términos de capacidad, compromiso o motivación de los participantes. Esto ocurrió cuando los elementos de gamificación fueron mal integrados o cuando el diseño no vinculó de manera efectiva los objetivos educativos con la experiencia del usuario (Yasin et al., 2025). Asimismo, los contenidos complejos y altamente técnicos requieren múltiples sesiones para garantizar una comprensión profunda (Yasin et al., 2019).

Por otro lado, se ha destacado la importancia de diseñar cuidadosamente los juegos, considerando la complejidad y la claridad de los objetivos, para lograr impactos significativos en el comportamiento. Se identificaron tres aspectos clave: los juegos como una forma de entretenimiento que genera relajación y motivación; los juegos como herramientas para mejorar habilidades cognitivas, facilitando la comprensión de temas complejos; y los juegos como facilitadores de cambios actitudinales, especialmente en contextos específicos (Mota et al., 2016). Dentro del diseño, ciertos elementos se mencionan con mayor frecuencia por

su capacidad para comprometer a los jugadores y promover cambios en su comportamiento. Entre ellos se encuentran que los desafíos, recompensas, puntuaciones y retroalimentación en tiempo real juegan un papel central (Hammady et al., 2022; Batzos et al., 2023). Asimismo, las dinámicas competitivas y la personalización aumentan la inmersión, la motivación y la relevancia personal del tema. Los sistemas de recompensas, como medallas y clasificaciones, permiten medir el progreso y mantener el interés, la satisfacción y el disfrute (Hammady et al., 2022; Boncu et al., 2022; Yildirim, 2017). Además, la competencia y la cooperación fomentaron una mayor participación en actividades grupales e individuales (Yildirim, 2017). La narrativa proporcionó un contexto inmersivo, mientras que los desafíos ofrecieron oportunidades para superar obstáculos, aumentando el interés y la motivación (Bitrián et al., 2023). Un desafío pendiente es diseñar mecanismos que mantengan el compromiso de los participantes a largo plazo (Boncu et al., 2022).

Finalmente, es importante destacar que el diseño de los juegos puede tener efectos diferenciados en hombres y mujeres (Egashira et al., 2022; Grevelink, 2015). Asimismo, no todos los participantes responden de la misma manera a las dinámicas de juegos. Aquellos que cuentan con una alta motivación intrínseca tienden a beneficiarse más, mientras que quienes suelen dependen de recompensas externas requieren más frecuentemente incentivos específicos para mantener su interés (Buckley et al., 2016). Además, los distintos elementos gamificados pueden generar efectos diversos según los rasgos de personalidad de los usuarios. Por ello, para garantizar el éxito de una intervención basada en juegos, es fundamental considerar sus características individuales e implementar enfoques personalizados (Smiderle et al., 2020).

4. Metodología

4.1 Teoría del Cambio: Hipótesis Causal

A continuación, se explica la hipótesis causal del estudio a través del enfoque de la **teoría del cambio**, una herramienta metodológica que explica cómo las actividades de una intervención generan una cadena de resultados que, a su vez, conducen al logro de los impactos finales deseados (Rogers, 2014). Esta metodología tiene como objetivo principal proporcionar claridad y coherencia en los procesos de diseño, planificación y evaluación, asegurando que las estrategias implementadas aborden de manera efectiva los problemas identificados. Además, este enfoque permite a los responsables de las intervenciones analizar de manera detallada las relaciones entre los insumos, actividades, resultados e impactos finales, estableciendo un camino lógico y verificable desde la situación inicial hasta la meta deseada (Glennerster y Takavarasha, 2013; Rogers, 2014; Bueno y Osuna, 2013).

Definición del Problema

Las campañas de ciberseguridad buscan sensibilizar a los usuarios sobre las amenazas digitales, fomentar comportamientos seguros, brindar las habilidades necesarias a través de la educación y la capacitación, así como influir en el cambio de actitudes y percepciones frente a la seguridad digital. No obstante, suelen fracasar en la promoción de conductas seguras debido a una combinación de factores humanos, así como a deficiencias en sus diseños (ver sección 3.1).

A la vez, el *phishing* es una técnica de fraude cibernético que **busca engañar a las personas** para que revelen información sensible, como credenciales de acceso o datos financieros, generalmente mediante correos electrónicos, mensajes o sitios web fraudulentos (Ferreira et al., 2015). En lugar de explotar vulnerabilidades técnicas en los sistemas, el *phishing* se basa en la **manipulación psicológica** de sus víctimas, utilizando principios de ingeniería social para inducirlas a realizar acciones que beneficien al atacante (Jari, 2022).

A nivel local, las principales fuentes de información para los consumidores sobre el *phishing* provienen de las campañas de seguridad promovidas por instituciones financieras y organismos reguladores. No obstante, al considerar tanto a proveedores financieros bancarios como no bancarios, no todos cuentan con campañas específicas contra el fraude en medios de pago online en sus sitios web institucionales. En los casos en que sí las ofrecen, estas suelen ser parciales, sin abordar todos los tipos de fraudes a los que están expuestos los usuarios. Además, la información disponible tiende a ser incompleta y de difícil acceso (SERNAC, 2025a).

De igual manera, no existen evaluaciones que midan la efectividad de estas campañas en la prevención de fraudes, especialmente en lo que respecta a técnicas de manipulación utilizadas para engañar a los consumidores. Sin embargo, extractos de reclamos ingresados a SERNAC evidencian que los usuarios pueden interpretar de manera restrictiva las recomendaciones entregadas. Un ejemplo de esto es la confusión en torno a conceptos como "clave de seguridad" y el alcance preciso de "compartir las claves" (SERNAC, 2025a).

Por otro lado, el análisis cuantitativo y cualitativo de los reclamos por fraude ingresados por consumidores al SERNAC durante 2023 (aproximadamente 10.000 casos) reveló una mayor participación de mujeres en todas las modalidades de fraude detectadas. En particular, se destaca su alta representación en los casos de *phishing*, donde 2 de cada 3 víctimas que presentan reclamos son mujeres (SERNAC, 2025a).

Del mismo modo, una encuesta realizada a 2.000 consumidores entre marzo y abril de 2024 confirma esta tendencia. Entre quienes declararon haber sido víctimas de fraude en línea, se estimó que las mujeres tienen un 4 % más de probabilidad de sufrir este tipo de delitos en comparación con los hombres. Por otro lado, contar con un nivel de experiencia alto o superior al promedio en banca en línea reduce en un 8 % la probabilidad de haber sido víctima de fraude (SERNAC, 2025b).

Ante este contexto, el presente estudio tiene como objetivo identificar y evaluar elementos comunicacionales que puedan ser más efectivos en las campañas contra el *phishing*. Se analiza cómo estos elementos contribuyen, por un lado, a capacitar a las personas en la detección de correos fraudulentos y, por otro, a influir positivamente en sus actitudes y percepciones sobre la seguridad digital. Tanto el fortalecimiento de esta capacidad como el cambio en actitudes y percepciones se consideran factores esenciales para generar modificaciones en el comportamiento observable (Petty er al, 2019).

Hipótesis Causal

Hipótesis causal: *Si los consumidores son expuestos a campañas contra el phishing con herramientas comunicacionales más persuasivas, estarán mejor preparados para identificar correos fraudulentos ("Experimento 1") y mostrarán una mayor disposición a adoptar, de manera preventiva, las recomendaciones de seguridad ("Experimento 2").*

Factores utilizados en los tratamientos**: Herramientas de comunicación más persuasivas**

Para reforzar una campaña estándar de detección de correos fraudulentos, se propusieron tres elementos comunicacionales:
Mensajes motivacionales ("Motivación"): Refuerzan la autoeficacia y resaltan la gravedad del phishing. Un estudio previo realizado por SERNAC (marzo-abril de 2024) concluyó que la autoeficacia es el factor más determinante para fomentar la autoprotección, seguido de la percepción de la gravedad del fraude (Sernac, 2025b).
Explicación de técnicas de ingeniería social ("Explicación"): Ofrece una descripción breve de los métodos de manipulación empleados en correos fraudulentos, facilitando así su identificación.
Enfoque lúdico ("Juego"): Consiste en un test para evaluar la legitimidad de los correos, lo que permite a los participantes practicar las recomendaciones previamente entregadas. Este método contribuiría a un proceso de persuasión comunicacional más efectivo y favorece la inmersión en la narrativa de la campaña**.**

Experimento 1:
En el Experimento 1 se analiza de qué manera la inclusión de *explicaciones* sobre las técnicas de ingeniería social empleadas en correos fraudulentos —como parte de una campaña estándar contra el *phishing*— y la incorporación de *mensajes motivacionales* que refuerzan la autoeficacia y la percepción de gravedad influyen tanto en la probabilidad de clasificar correctamente un correo (ya sea fraudulento o legítimo) como en la cantidad de señales de *phishing* que se reconocen y utilizan para fundamentar dicha clasificación.

Variables dependientes:

- Probabilidad de diferenciar correctamente entre correos fraudulentos y legítimos ("*probabilidad*").
- Número de elementos identificados como señales de *phishing* ("*N°elementos*") utilizados para clasificar los correos como fraudulentos o legítimos.

El experimento se formuló bajo las siguientes hipótesis de investigación:

H1: La *probabilidad* de clasificar correctamente un correo, ya sea legítimo o fraudulento, aumenta cuando la campaña de concienciación incluye una *explicación* sobre las técnicas de ingeniería social utilizadas en los correos, en comparación con una campaña sin esta información.

H2: La *probabilidad* de clasificar correctamente un correo, ya sea legítimo o fraudulento, aumenta cuando la campaña incluye *mensajes motivacionales* que refuerzan la autoeficacia y la percepción de gravedad, en comparación con una campaña que no los incorpora.

H3: Se identifica una mayor cantidad de elementos que permiten clasificar correctamente un correo cuando la campaña incluye una explicación sobre técnicas de ingeniería social, en comparación con una campaña que no las incorpora.

H4: Se identifica una mayor cantidad de elementos que permiten clasificar correctamente un correo cuando la campaña proporciona mensajes motivacionales que refuerzan la autoeficacia y la percepción de gravedad de las estafas de *phishing*, en comparación con una campaña que no los incorpora.

Experimento 2:

El Experimento 2 analiza cómo distintos elementos comunicacionales, incluidos en las campañas estándar de *phishing*, influyen en los cambios de actitud hacia las medidas de detección de fraudes en línea (*actitud*) y en la intención de protegerse (*intención*). Estas variables se consideran esenciales para fomentar modificaciones en el comportamiento observable, favoreciendo una mayor disposición a adoptar preventivamente las recomendaciones de seguridad.

Los elementos comunicacionales evaluados incluyen: a) *Explicaciones* sobre técnicas de ingeniería social, b) los *mensajes motivacionales* que apelan la autoeficacia y la percepción de gravedad y c) el *enfoque de juegos*, que permite la inmersión completa de los usuarios en la narrativa del mensaje de la campaña.

El experimento se formuló bajo las siguientes hipótesis de investigación:

Hipótesis sobre Cambios en la Actitud hacia las medidas de seguridad e Intención de Protección:

H5: Las personas desarrollan una mejor *actitud* hacia las medidas de seguridad cuando las campañas *explican* las técnicas de ingeniería social utilizadas en los correos fraudulentos, en comparación con aquellas que no lo hacen.

H6: Las personas desarrollan una mejor *actitud* hacia las medidas de seguridad cuando las campañas incluyen *mensajes motivacionales* que refuerzan la autoeficacia y la percepción de gravedad, en comparación con aquellas que no lo hacen.

H7: Las personas desarrollan una mejor *actitud* hacia las medidas de seguridad cuando las campañas incluyen *elementos de juego*, en comparación con aquellas que no lo hacen.

H8: La inclusión de *explicaciones* sobre las técnicas de ingeniería social en las campañas de *phishing* aumenta la *intención* de las personas de revisar señales de estafa en correos electrónicos antes de responderlos, en comparación con campañas sin esta información.

H9: La inclusión de *mensajes motivacionales* que refuerzan la autoeficacia y la percepción de gravedad en las campañas de *phishing* aumenta la *intención* de las personas de revisar señales de estafa en correos electrónicos antes de responderlos, en comparación con campañas que no contienen este elemento.

H10: La inclusión de un *enfoque de juego* en las campañas de *phishing* aumenta la *intención* de las personas de revisar señales de estafa en correos electrónicos antes de responderlos, en comparación con campañas que no utilizan este enfoque.

Hipótesis sobre Cambios en variables asociadas al Canal de Persuasión:

H11: La *Atención* hacia la campaña mejora cuando las campañas son reforzadas con las herramientas comunicacionales propuestas.

H12: La *Comprensión* sobre *phishing* es mayor en personas expuestas a campañas reforzadas con las herramientas comunicacionales propuestas.

H13: La *aceptación general de la campaña* (medida en términos de *Gusto General, Efectividad Percibida y Fuerza Argumental Percibida*) es mayor cuando las campañas son reforzadas con las herramientas comunicacionales propuestas.

H14: La *percepción de autoeficacia* en la detección de correos fraudulentos es mayor en personas expuestas a campañas reforzadas.

H15: La *percepción de autoeficacia* en la detección de correos fraudulentos es mayor en personas expuestas a campañas reforzadas.

Resultados esperados:

a) Mayor probabilidad de que las personas distingan correos fraudulentos de correos legítimos, detectando un mayor número de señales de manipulación.
b) Aumentar la Motivación de Protección, reflejado en una mayor autoeficacia y gravedad percibida.
c) Mejor desempeño de las variables asociadas al Modelo de Persuasión, evidenciado en mayores niveles de Atención, Comprensión y Aceptación General de la campaña.
d) Mayor Aceptación de la campaña, en términos de Gusto, Efectividad Percibida y Fuerza Argumental Percibida.
e) Mejora en la Actitud hacia las medidas de detección de fraudes en línea.
f) Incremento en la Intención de protegerse en el futuro mediante la revisión de señales de estafa en los correos antes de responder (Intención).

4.2 Método Experimental

4.2.1 Estrategia de Identificación

Como enfoque metodológico, este estudio adopta los Ensayos Controlados Aleatorizados (RCT, por sus siglas en inglés), un diseño experimental que permite establecer relaciones causales entre las variables de interés. Para el análisis de los resultados, se emplearon dos enfoques estadísticos. Primero, se evaluaron los efectos a nivel de tratamientos mediante el ***Cell Means Model***. Luego, se aplicó un **ANOVA factorial**, que permite analizar el impacto de distintos factores y sus interacciones en la variable de respuesta. Dado que ambos enfoques requieren ciertos supuestos sobre la distribución de los datos, como medida de robustez también se incorporaron modelos no paramétricos, como la prueba de **Kruskal-Wallis**. A continuación, se presentan cada uno de estos métodos.

4.2.1.1 Diseño Experimental: Ensayos controlados Aleatorizados

Los ensayos controlados aleatorizados, conocidos como ***Randomized Controlled Trials* (RCT),** son una metodología experimental utilizada para evaluar el impacto causal de una intervención o tratamiento mediante la comparación de grupos: uno que recibe la intervención y otro que actúa como control. Esta metodología es considerada el gold standard en investigación experimental, ya que permite minimizar sesgos y establecer relaciones causales con mayor rigor.

Dado que el efecto de un tratamiento se define como la diferencia en la variable de resultado de un individuo con y sin tratamiento—una situación inobservable, ya que un mismo individuo no puede estar simultáneamente en ambas condiciones—es necesario emplear un grupo de comparación. Así, el contrafactual se estima comparando los resultados del grupo tratado con los del grupo de control, el cual es equivalente en todas sus características observables y no observables, excepto en la exposición al tratamiento (Gertler et al., 2016, Serin et al., 2022).

La aleatorización de los participantes asegura que cualquier diferencia observada entre los grupos pueda atribuirse exclusivamente a la intervención y no a factores externos (Duflo et al., 2006). Su objetivo principal es generar grupos estadísticamente equivalentes, equilibrados tanto en características observables como no observables (Torres, 2021). Gracias a esto, los RCT eliminan el sesgo de selección, es decir, las diferencias previas en variables no observadas entre los grupos de tratamiento y control, lo que facilita una estimación más precisa del efecto causal de la intervención.

4.2.1.2 Modelo basado en tratamientos

El Modelo Basado en Tratamientos (*Cell Means Model*) es un enfoque estadístico utilizado en el análisis de datos experimentales, especialmente en el contexto de ensayos controlados aleatorizados (RCT, por sus siglas en inglés). Su principal característica es que estima y compara directamente las medias de cada grupo sin descomponer los efectos en términos de factores individuales o interacciones. Este modelo resulta útil cuando el interés principal es evaluar el impacto de diferentes condiciones experimentales, considerándolas como unidades independientes sin necesidad de asumir una estructura factorial subyacente.

A la vez, para medir el impacto de la intervención en estos modelos, se comparan los promedios de un indicador específico entre el grupo tratado y el grupo de control. Cualquier diferencia en los resultados entre ambos grupos refleja el **efecto promedio del tratamiento** *(Average Treatment Effect* – ATE, en inglés) (Gertler et al., 2016).

En el presente experimento, se busca identificar el **impacto o efecto causal promedio** de cada tratamiento, entendido como una combinación específica de factores (Motivación y Explicación y Juego). La siguiente ecuación representa el modelo utilizado en el Experimento 2, en el cual el grupo control (T_5) está conformado por los participantes que recibieron la campaña base, en la que todos los factores están ausentes. Dado que el grupo de control se excluye de la ecuación, su efecto queda representado por el intercepto del modelo (β_0). Por su parte, cada coeficiente en la ecuación (β_i) cuantifica el **impacto causal de cada tratamiento** (T_i), expresado como la diferencia entre el resultado del grupo que recibió dicho tratamiento y el del grupo control.

El modelo de regresión es el siguiente:

$$y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 T_8 + \varepsilon$$

Donde cada variable, T_i, es una *dummy* que toma el valor de 1 si se presenta una combinación específica de factores.

Dado que la mayoría de las variables dependientes en este estudio se construyen como el promedio de otras variables medidas en una escala de Likert, pueden tratarse como continuas o, alternativamente, como ordinales. En cada caso, el método de inferencia estadística más adecuado será diferente. Existe un amplio debate sobre la conveniencia de emplear pruebas estadísticas paramétricas para datos recopilados en escalas de Likert (Kraska-Miller, 2013). En consecuencia, este estudio aplica tanto pruebas paramétricas como no paramétricas para corroborar la robustez de las inferencias realizadas.

Si las variables se consideran **continuas,** se aplica la **prueba t de Student** para comparar medias entre grupos. Aunque esta prueba asume normalidad en la distribución de la variable dependiente dentro de cada grupo, sigue siendo válida cuando el tamaño muestral es lo suficientemente grande para que se aplique el teorema del límite central. Además, si se emplea un modelo de regresión, se pueden estimar errores estándar robustos a la heterocedasticidad para mejorar la precisión de las inferencias.

Otro aspecto clave a considerar es que, al evaluar la significancia de múltiples coeficientes mediante pruebas t de Student independientes, se incrementa la probabilidad global de cometer un error Tipo I, un fenómeno conocido como el problema de comparaciones múltiples. Es decir, a medida que aumenta el número de pruebas, también lo hace la probabilidad de obtener falsos positivos, identificando efectos significativos que en realidad no existen (Cady, 2017). Para mitigar este problema, una alternativa es aplicar correcciones para comparaciones múltiples, como el ajuste de Bonferroni, Holm o FDR *(False Discovery Rate*) (Montgomery, 2010; Oehlert, 2000).

Por otro lado, si las variables se consideran ordinales, es recomendable utilizar pruebas no paramétricas, ya que no asumen normalidad ni intervalos equidistantes entre categorías. Un ejemplo es la **Prueba de Kruskal-Wallis**, que es una alternativa no paramétrica al ANOVA de una vía para comparar más de dos grupos. Esta prueba analiza si existen diferencias en la distribución de los valores entre los grupos basándose en los rangos de los datos. Sin embargo, si el resultado indica una diferencia significativa, no señala qué grupos difieren entre sí. Para ello, se utiliza la prueba de Dunn, una comparación post hoc que permite identificar qué grupos presentan diferencias significativas, ajustando por Bonferroni los valores p para evitar errores tipo I. En términos generales, la prueba de Kruskal-Wallis evalúa si las distribuciones de los grupos difieren en su ubicación o forma. Si las distribuciones de los grupos tienen formas y dispersiones similares, el test puede interpretarse como una

comparación de medianas. Pero si las distribuciones son asimétricas o tienen dispersión diferente, no necesariamente está comparando solo medianas (Kraska-Miller, 2013).

4.2.1.3 Modelo basado en Factores

El Análisis de Varianza Factorial (ANOVA Factorial) es una técnica estadística utilizada para analizar el efecto de múltiples factores sobre una variable dependiente, permitiendo evaluar tanto sus efectos individuales como sus interacciones. Se utiliza ampliamente en estudios experimentales, incluyendo ensayos controlados aleatorizados. Esta herramienta es fundamental para analizar la variabilidad en un conjunto de datos y descomponerla en componentes atribuibles a los factores experimentales y al error aleatorio (Field, 2018; Montgomery, 2019).

El **ANOVA factorial** permite evaluar simultáneamente la influencia de dos o más variables independientes categóricas (factores) sobre una variable dependiente continua. A la vez, este enfoque facilita el estudio de los efectos principales e interacciones dentro de un mismo modelo, evitando la necesidad de realizar múltiples pruebas independientes y reduciendo así el riesgo global de error Tipo I (Larson-Hall, 2015; Field, 2018; Tabachnick & Fidell, 2013). Es particularmente útil en investigaciones que buscan identificar tanto el impacto individual de cada factor como la posible interacción entre ellos (Montgomery, 2019).

Para que los resultados del ANOVA factorial sean válidos, deben cumplirse ciertos supuestos estadísticos (Field, 2018; Montgomery, 2019; Tabachnick & Fidell, 2013): a) Independencia de las observaciones: las mediciones deben ser independientes dentro y entre los grupos experimentales, evitando estructuras de dependencia que puedan invalidar las pruebas de significancia; b) Normalidad de los residuos: Los errores del modelo deben seguir una distribución normal; c) Homogeneidad de varianzas (homocedasticidad): la varianza de los residuos debe ser homogénea en todos los niveles de los factores, ya que la heterocedasticidad puede afectar la robustez de los resultados; d) Especificación adecuada del modelo: el modelo debe incluir los efectos principales y las interacciones necesarias para una correcta representación de los efectos factoriales, evitando la omisión de términos relevantes que puedan generar sesgo en la interpretación de los resultados.

En el contexto experimental de este estudio, el análisis del efecto de los factores busca identificar la influencia de factores como *Motivación, Explicación* y *Juego*, considerando sus posibles interacciones. Esto implica que el impacto de un factor puede variar en función de la presencia o interacción con otro.

El modelo de regresión es el siguiente:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3 + e$$

En este modelo, se consideran tres factores: *Explicación, Motivación y Juego*, representados respectivamente por x_1, x_2 y x_3 . Estos factores se codifican como variables dummy, asignándoles un valor de 1 cuando el factor está presente y 0 cuando está ausente. El efecto de cada factor puede depender de la presencia o ausencia de los demás, ya que cada uno tiene dos niveles (presente y ausente). Los ***efectos principales*** se miden a través de los coeficientes β_1, β_2, β_3, representan el impacto individual de cada factor sobre la variable dependiente. Por otro lado, los ***efectos de interacción*** se capturan mediante β_4, β_5, β_6, β_7, reflejando cómo la combinación de factores influye en la respuesta. Además, en un ANOVA factorial con codificación *dummy*, la constante del modelo (β_0) representa el valor esperado de la variable dependiente cuando todos los factores están ausentes, es decir, cuando todas

las variables *dummies* toman el valor cero, o dicho de otro modo, representa la media del grupo de referencia (Langsrud, 2003).

Además, en un modelo ANOVA, la inferencia se basa en la Prueba F, que evalúa si los factores y sus interacciones explican una variabilidad significativa en la variable dependiente. La prueba compara la variabilidad explicada por cada factor con la variabilidad residual, determinando si la variabilidad explicada por un factor es significativamente mayor que la variabilidad dentro de los grupos. Un *p-valor* significativo en la prueba F (por ejemplo, $p<0.05$) indica que un factor influye significativamente en la variable dependiente, mientras que un valor no significativo sugiere que las diferencias observadas podrían deberse al azar y no a los efectos de los factores. Además, la prueba F global evalúa si al menos uno de los factores tiene un efecto significativo sobre la variable dependiente, comparando la variabilidad explicada por el modelo con la variabilidad residual (Langsrud, 2003).

Dado el tipo de datos y el diseño experimental, además del análisis tradicional del ANOVA factorial, se llevó a cabo un **ANOVA factorial Tipo III,** como prueba de robustez adicional. Este enfoque es particularmente útil en contextos con desequilibrios en los datos. En este caso, existe una leve distribución desigual de participantes entre los niveles de los factores que podría requerir una metodología que ajuste los efectos considerando estos desbalances (Lawson, 2014).

Este enfoque facilita la interpretación de los efectos de los factores, ya que permite comparar cada nivel con la media global en lugar de con un grupo de referencia arbitrario (Lawson, 2014, Langsrud, 2003).

El ANOVA de tipo III permite realizar ajustes mediante contrastes de suma a cero (sum contrasts), lo que posibilita que el modelo capture las diferencias relativas sin depender del número absoluto de observaciones en cada nivel o factor. En este enfoque, los efectos principales y las interacciones se estiman controlando por todos los demás términos del modelo, asegurando que los coeficientes reflejen efectos marginales ajustados. Para lograr esto, las variables categóricas (factores) se codifican de manera que la suma de los coeficientes asignados a sus distintos niveles sea igual a cero. Esto garantiza que el intercepto del modelo represente la media global de la variable dependiente (promedio considerando todas las observaciones, independientemente de los niveles de los factores) y que cada coeficiente refleje la desviación de su nivel con respecto a dicha media. Este método podría facilitar la interpretación de los efectos factoriales, ya que permite comparar cada nivel con la media global en lugar de con un grupo de referencia arbitrario (Lawson, 2014; Langsrud, 2003).

4.2.1.4 Tamaños de los Efectos

En estudios experimentales, el **tamaño del efecto** es una medida estadística que cuantifica la magnitud de la diferencia entre grupos o el grado de asociación entre variables. A diferencia de las pruebas de hipótesis de significación estadística (p-valores), que indican si hay suficiente evidencia para rechazar la hipótesis nula (la cual plantea que no hay diferencia entre grupos o que las variables no están asociadas), el tamaño del efecto proporciona información sobre su relevancia práctica (Kelley & Preacher, 2012).

El reporte del tamaño del efecto es esencial, ya que permite evaluar la magnitud del impacto de una intervención y facilita la comparación entre estudios, especialmente en análisis combinados como los meta-análisis. Además, desempeña un papel clave en el cálculo del poder estadístico, que mide la probabilidad de rechazar correctamente la hipótesis nula cuando en realidad es falsa, es decir, la probabilidad de encontrar un efecto significativo si

este realmente existe en la población, reduciendo así la probabilidad de cometer errores Tipo II (falsos negativos). Su correcta estimación resulta fundamental para determinar el tamaño muestral necesario en futuros experimentos.

En este estudio se utilizaron diferentes medidas del tamaño del efecto, cada una adecuada para situaciones estadísticas diferentes: (1) *d* de Cohen, (2) Eta cuadrada (η^2), (3) Eta cuadrada parcial (η_p^2), y (4) *r* de Cohen (*r*). Estas medidas se describen a continuación.

Modelo basado en tratamientos

La ***d* de Cohen** (Cohen, 1988) es una medida del tamaño del efecto que representa la diferencia entre las medias de dos grupos en unidades de desviación estándar.

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

Cohen's d se utiliza principalmente en comparaciones de dos grupos (como en pruebas t de Student para muestras independientes, utilizadas en estudios experimentales donde se quiere cuantificar la magnitud de la diferencia entre un grupo de control y un grupo experimental). Un valor de d = 0 significa que las medias de los dos grupos son iguales (ningún efecto). Valores mayores en magnitud indican diferencias más grandes entre grupos. Cohen (1988) propuso convenciones para interpretar d (aunque son pautas generales y el contexto específico siempre importa): valores menores 0,20 se considera un efecto muy pequeño (d=0.2 implica que las medias difieren en una quinta parte de la desviación estándar), entre 0.2 y 0.5 es un efecto pequeño, entre 0.5 y 0.8 es un efecto mediano y sobre 0,80 es un efecto grande (Cohen, 1988).

La d de Cohen se usa idealmente con datos continuos (o variables ordinales si se asumen intervalos equivalentes), normalmente distribuidos, con varianza homogénea e independencia.

Modelo basado en Factores

Eta cuadrada (η^2) es una medida de tamaño del efecto que representa la proporción de varianza total de la variable dependiente explicada por un factor o variable independiente. En un análisis de varianza (ANOVA), η^2 se calcula como la suma de cuadrados del efecto dividida por la suma de cuadrados total:

$$\eta^2 = \frac{SS_{effect}}{SS_{total}}$$

Su valor oscila entre 0 y 1 (0% a 100%) indicando qué porcentaje de la variabilidad total en los datos se debe al efecto estudiado. Por ejemplo, un η^2=0.10 sugiere que el 10% de la varianza total en la variable de resultados es atribuible al factor en cuestión.

En el ANOVA factorial a menudo se reporta el **eta cuadrado parcial (η_p^2)** para cada factor, que es una medida similar, en la que se eliminan los efectos de otras variables independientes e interacciones (Richardson, 2011).

$$\eta_p^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$$

Donde *SS* denota la *suma de cuadrados*, que puede ser la variabilidad total en los datos (SS_{total}), aquella explicada por el tratamiento o factor (SS_{effect}), o la varianza residual (no explicada) de la estimación (SS_{error}). Requiere usar variables continuas, distribuidas normalmente, con varianza homocedástica e independencia de las observaciones de cada grupo.

Los valores se interpretan de la siguiente manera: $\eta_p^2 = 0.01$ es un efecto pequeño; $\eta_p^2 = 0.06$ es un efecto mediano; $\eta_p^2 = 0.14$ es un efecto grande.

Es importante destacar que, al ser una fracción de varianza, η^2 tiende a valores más bajos en comparación con medidas como d de Cohen; incluso un porcentaje aparentemente pequeño puede ser significativo dependiendo del contexto (p. ej., en estudios con muchos factores que afectan el resultado).

Modelo no Paramétrico

La prueba de Kruskal-Wallis es una alternativa no paramétrica al ANOVA de una vía para comparar k grupos independientes. Sin embargo, Kruskal-Wallis solo indica si existen diferencias globales entre los grupos, sin identificar qué grupos difieren específicamente. Para ello, se realizan comparaciones post hoc, y una prueba comúnmente utilizada es la prueba de Dunn, que permite realizar comparaciones múltiples con ajuste para controlar el error tipo I (por ejemplo, con las correcciones de Bonferroni o Holm) (Dunn, 1964).

Para medir el tamaño del efecto en pruebas no paramétricas, se usa **la correlación biserial de rango** (**r de Rank-Biserial),** que representa la relación entre dos grupos basada en rangos. Su cálculo para una comparación entre dos grupos se expresa como:

$$r = \frac{|Z|}{\sqrt{N}}$$

Donde Z es el estadístico de la prueba de Dunn y N = $n_1 + n_2$ es el tamaño total de observaciones de los dos grupos involucrados en esa comparación específica.
La interpretación de r sigue los criterios estándar de magnitud: 0.1 pequeño, 0.3 mediano, 0.5 grande. Esto permite comparar la fuerza de las diferencias entre distintas parejas de grupos.

Esta medida es especialmente adecuada para datos ordinales o no normales, ya que, a diferencia de medidas como d de Cohen (que se basa en medias y desviaciones estándar), r se fundamenta en los rangos y no requiere asumir normalidad en la distribución de los datos. Por ello, es una opción apropiada cuando la prueba estadística empleada es no paramétrica, proporcionando una cuantificación compatible del tamaño del efecto.

4.2.2 Diseño de la Campaña de *Phishing* Estándar

Las campañas de concientización sobre *phishing* promovidas por las instituciones financieras locales suelen ser muy heterogéneas, especialmente en su contenido. En general, estas campañas pueden incluir algunos de los siguientes elementos: definición del *phishing*, modus operandi, consecuencias, señales para identificar correos fraudulentos, recomendaciones ante una posible exposición, acciones a tomar en caso de haber sido víctima, medidas preventivas y ejemplos ilustrativos (Sernac, 2025a).

Para los fines de este estudio, la campaña evaluada **se centró únicamente en destacar un conjunto de elementos comunes presentes en los correos electrónicos que permiten detectar una estafa de *phishing*. Entre estos elementos se incluyen técnicas de ingeniería social utilizadas para engañar a los usuarios (Figura 1).**

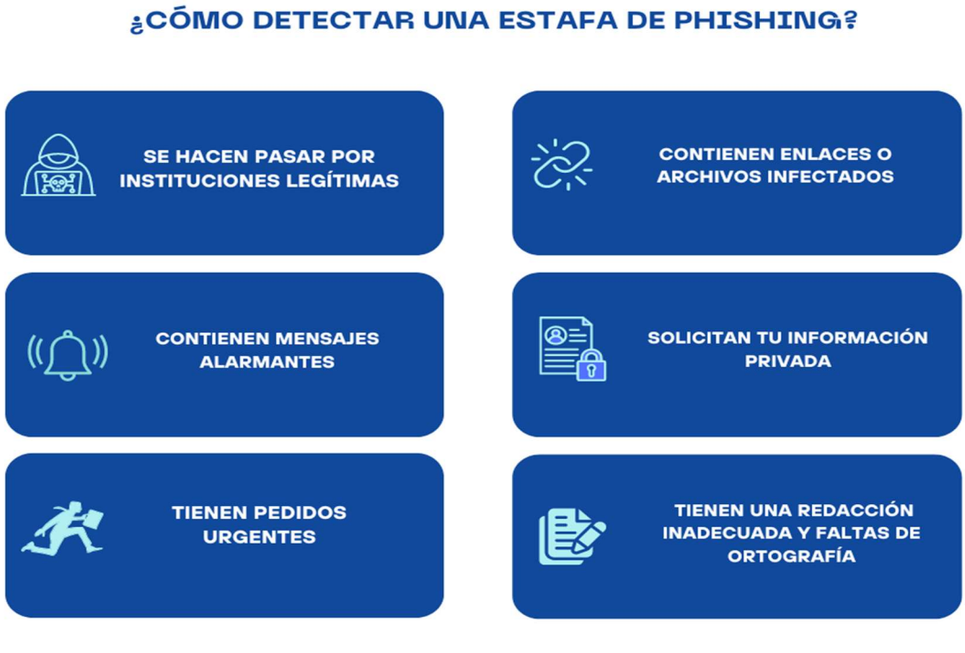
Asimismo, la campaña de *phishing* diseñada para el estudio debía presentar un diseño familiar y reconocible para los participantes, alineado con el estilo de comunicación utilizado por las instituciones financieras locales en sus distintos canales. Para definir su diseño, se realizó un análisis de los formatos empleados en las campañas de los bancos locales, tanto en sus páginas web como en los correos electrónicos enviados a sus clientes. En general, las campañas en las páginas web se caracterizaban por un contenido mayormente textual, con abundante información escrita y un uso limitado de imágenes. En contraste, los correos electrónicos estructuraban la información mediante mapas conceptuales, con textos resumidos y un diseño más visual, incorporando colores y elementos gráficos específicos para mejorar la presentación del mensaje.

En este contexto, se optó por un formato similar al utilizado en las campañas enviadas por correo electrónico replicando, en parte, el uso de formas, colores e íconos característicos de estas comunicaciones.

A la vez, el diseño experimental tomó en cuenta que en un diseño factorial, es fundamental que todas las condiciones experimentales sean equivalentes en formato, duración y contexto, excepto por las manipulaciones (factores) que se están evaluando. Por ejemplo, si se comparan dos tipos de mensajes, ambos deben mantener características idénticas, variando únicamente el elemento de interés. Esto permite aislar el efecto del factor y minimizar la influencia de variables externas en la atención. Un diseño riguroso debe incluir grupos de control o comparativos para diferenciar con precisión el impacto del tratamiento del "ruido de fondo" o de posibles factores de confusión (Montgomery, 2017; Shadish et al., 2002).

A pesar de ser un experimento de laboratorio, se procuró maximizar la naturalidad dentro de un entorno controlado. Para ello, se mantuvo a los participantes sin conocimiento del objetivo específico del estudio, evitando así respuestas sesgadas por efecto de demanda. Además, se reprodujeron contextos realistas de uso de correo electrónico y se controlaron rigurosamente las condiciones técnicas y sociales. Esto permitió obtener datos de comportamiento lo más cercanos posible a la realidad, sin comprometer la precisión y repetibilidad que proporciona el entorno de laboratorio.

Figura 1: Campaña de *Phishing* Estándar (Sin refuerzo)



Las campañas se optimizaron para dispositivos móviles mediante un diseño en columna, asegurando una correcta visualización del contenido. Aunque los participantes debían desplazarse verticalmente para completar la encuesta, que fue estructurada en bloques, no era necesario un desplazamiento horizontal ni ajustes manuales de zoom para una lectura clara en el celular, lo que minimizó el riesgo de sesgos en las respuestas. Estudios señalan que la falta de optimización en móviles incrementa la tasa de abandono, especialmente en pantallas pequeñas, dificultando la recolección de datos (Clement et al., 2020; Callegaro, 2010).

4.2.3 Diseño de los Factores

a) Diseño Mensaje Motivacional ("Motivación")

Los elementos motivacionales suelen estar presentes en grandes campañas de ciberseguridad promovidas por organismos internacionales y gobiernos, ya sea a través de eslogan o mensajes motivacionales. Un eslogan es una frase breve y memorable que resume la idea central de la campaña. Su función principal es captar rápidamente la atención y reforzar el mensaje clave de manera clara y efectiva. Por su parte, el mensaje motivacional amplía el significado del eslogan, buscando inspirar o persuadir a la audiencia para adoptar una actitud o comportamiento específico.

Por ejemplo, en Estados Unidos, la campaña con el eslogan **"Stop. Think. Connect."**, lanzada en el 2010 por el Departamento de Seguridad Nacional, promueve la reflexión antes de conectarse o compartir información en línea, enfatizando la importancia del pensamiento crítico para prevenir incidentes de ciberseguridad[1]. De manera similar, la Agencia de la Unión Europea para la Ciberseguridad (ENISA) lanzó la campaña el 2021 con el eslogan **"Think Before U Click"**, con el objetivo de concientizar a los usuarios sobre la importancia de detenerse y evaluar la legitimidad de correos electrónicos y enlaces antes de interactuar con ellos, reduciendo así el riesgo de comprometer su seguridad digital[2]. Por su parte, en el Reino Unido, la campaña con el slogan **"Take Five to Stop Fraud"** busca ayudar a las personas

[1] https://sherloc.unodc.org/
[2] https://digital-strategy.ec.europa.eu/en/news/think-u-click-european-cybersecurity-month-2020

a protegerse contra fraudes financieros prevenibles, especialmente aquellos en los que los delincuentes se hacen pasar por organizaciones de confianza[3].

Por otro lado, las instituciones privadas recurren con mayor frecuencia a mensajes motivacionales, como "Protégete, evitar el fraude está en tus manos" (España) [4] o "¡Conviértete en Súper Héroe! Actuemos juntos frente a las ciberamenazas" (Chile) [5]. Estos mensajes buscan principalmente motivar a las personas reforzando su **autoeficacia**, es decir, la creencia en las propias capacidades para organizar y ejecutar las acciones necesarias para lograr determinados objetivos o resultados (concepto desarrollado por Bandura,1977). A la vez, la evidencia empírica indica que una de las razones por las cuales las campañas suelen fracasar es que las personas tienden a ignorarlas (Caputo et al., 2014). Por ello, una estrategia efectiva podría ser sensibilizar al público mediante la exposición de las graves consecuencias de ser víctima de fraudes en línea.

Para definir el mensaje motivacional de este estudio, se consideraron la Teoría de Motivación de Protección (PMT) y el Modelo de Proceso Paralelo Extendido (EPPM).

La PMT explica cómo las personas responden a situaciones de riesgo mediante dos evaluaciones clave: la amenaza percibida y la capacidad de afrontamiento. Si un individuo se percibe vulnerable y considera la amenaza grave, es más probable que tome medidas preventivas. Además, si confía en la eficacia de la acción recomendada y en su capacidad para ejecutarla (autoeficacia), es más propenso a adoptar un comportamiento protector. Por su parte, el EPPM destaca el papel del miedo en la motivación de protección, señalando que un equilibrio adecuado entre la percepción de la amenaza y la autoeficacia es crucial para fomentar respuestas adaptativas, evitando generar un miedo excesivo que desincentive la acción.

Entre marzo y abril de 2024, un estudio en Chile con 2.000 consumidores analizó la aplicación de la PMT en el contexto del fraude financiero en línea. Los resultados indicaron que la autoeficacia es el factor más determinante para la protección, seguido de la percepción de la gravedad del fraude. Esto sugiere que una comunicación eficaz sobre los riesgos y consecuencias del fraude puede impulsar comportamientos protectores.

Con base en estos hallazgos, el **mensaje motivacional propuesto** enfatiza tanto la gravedad de la amenaza como la autoeficacia del usuario **(Figura 2).** Para su selección, se evaluaron varias frases, eligiendo aquellas con mayor retención en la memoria tras una hora de exposición.

Figura 2: Mensaje Motivación que apela a la gravedad y autoeficacia percibida

[3] https://www.takefive-stopfraud.org.uk/
[4] https://www.infobae.com/espana/agencias/2024/04/24/bancos-y-fuerzas-de-seguridad-impulsan-una-campana-para-evitar-los-fraudes-digitales/
[5] https://ciberseguridad.informatica.uc.cl/campanas/

b) Diseño Explicación de las Técnicas de ingeniería social presente en los correos ("Explicación")

El *phishing* suele ser peligroso y efectivo debido al uso de técnicas de manipulación psicológica, conocidas como ingeniería social, que se utilizan para engañar a sus víctimas. A diferencia de otros tipos de fraude, el *phishing* logra que la víctima participe activamente en el engaño (Jari, 2022).

La **ingeniería social** es el conjunto de técnicas que explotan la predisposición humana a la confianza, la obediencia a la autoridad y la conformidad social, entre otros factores psicológicos (Cialdini, 2007). Los ataques de *phishing* se apoyan en estos principios para generar escenarios convincentes que reduzcan el escepticismo y lleven a la víctima a actuar de manera impulsiva o automática (Gragg, 2003). Algunos de los métodos más comunes incluyen la suplantación de identidad de figuras de autoridad, la inducción de emociones fuertes como miedo o urgencia, y la creación de relaciones falsas para obtener información privilegiada (Stajano & Wilson, 2011). Además, investigaciones recientes han señalado que las técnicas de ingeniería social aplicadas en el *phishing* pueden ser incluso más efectivas cuando combinan múltiples principios psicológicos simultáneamente, aumentando la probabilidad de éxito del ataque (Albladi & Weir, 2018).

A continuación, se presentan las principales estrategias de ingeniería social utilizadas en los ataques de *phishing*, explicando cómo los atacantes las emplean para manipular a sus víctimas y maximizar su efectividad (Stajano y Wilson, 2011; Gragg, 2003; y Cialdini, 2007):

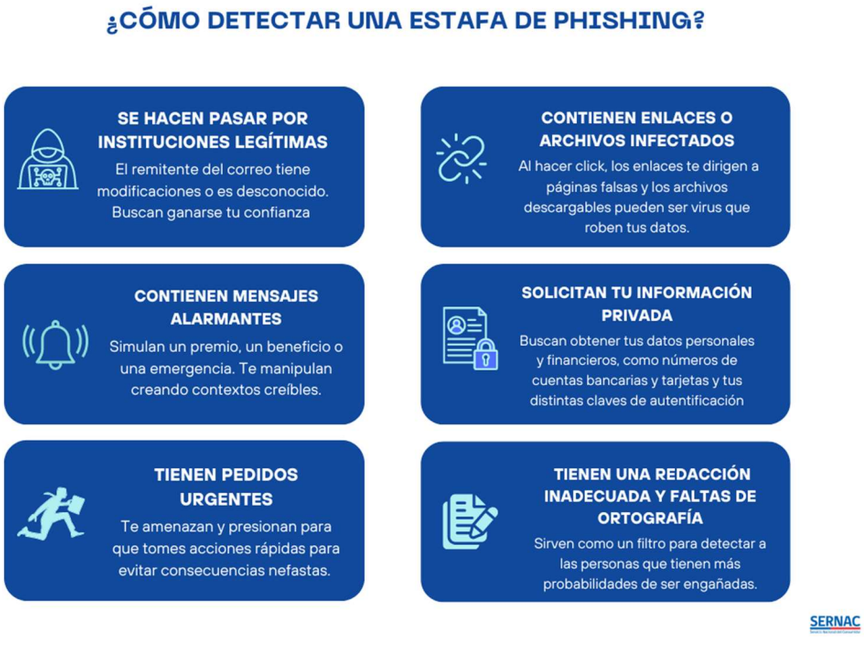
- **Autoridad:** Las personas tienden a aceptar instrucciones de figuras de autoridad sin cuestionarlas. En el contexto del *phishing*, los atacantes se hacen pasar por representantes de instituciones confiables para lograr que sus víctimas sigan sus indicaciones sin sospechar.
- **Prueba social/ Rebaño:** El comportamiento humano se ve influenciado por la tendencia a seguir lo que hacen los demás, especialmente cuando los riesgos son compartidos. Los estafadores se aprovechan de esta inclinación presentando pruebas falsas de que otras personas han tomado la misma acción, haciendo que la víctima perciba la petición como legítima.
- **Agrado/Similitud:** Las personas suelen confiar más en aquellos con quienes comparten intereses o afinidades. Los atacantes también pueden crear identidades falsas que copian las preferencias y valores de la víctima con el fin de establecer un vínculo de confianza y facilitar el engaño.
- **Compromiso/ Consistencia**: Una vez que una persona se ha comprometido con una acción, es más propensa a mantenerse coherente con esta. En el *phishing*, los estafadores pueden empezar con peticiones pequeñas para generar un sentido de compromiso, conduciendo gradualmente a la víctima a realizar acciones más significativas.
- **Escasez:** La percepción de que un recurso es limitado genera urgencia y una respuesta emocional impulsiva. A sabiendas de esto, los delincuentes suelen utilizar mensajes que advierten sobre la posible pérdida de acceso o de una oportunidad exclusiva, forzando a la víctima a actuar apresuradamente sin analizar la situación.
- **Reciprocidad:** Las personas sienten la necesidad de devolver favores recibidos. Los estafadores pueden ofrecer ayuda o beneficios aparentes con la intención de generar en la víctima un sentido de obligación que la lleve a cumplir sus solicitudes.
- **Inducción de emociones fuertes:** Las emociones intensas, como el miedo o la excitación, pueden afectar el pensamiento crítico. Los ciberdelincuentes emplean tácticas que generan pánico o sensación de urgencia para inducir respuestas impulsivas sin un análisis racional previo.

- **Sobrecarga:** Exponer a una persona a un exceso de información o a múltiples peticiones simultáneas puede saturar su capacidad de procesamiento. En el *phishing*, esta táctica se usa para hacer que la víctima acepte instrucciones sin examinar detenidamente su validez.
- **Relaciones engañosas**: Los atacantes también crean relaciones falsas con sus víctimas, compartiendo información o intereses en común para desarrollar confianza y manipularlas emocionalmente con el fin de obtener datos confidenciales o inducir acciones perjudiciales para las mismas.
- **Difusión de Responsabilidad y Deber Moral**: Cuando las personas sienten que su responsabilidad está compartida con otros o que están cumpliendo un deber moral, es más probable que realicen acciones que de otro modo dudarían de ejecutar. Los defraudadores pueden explotar este fenómeno para minimizar la resistencia de la víctima a ejecutar lo solicitado.
- **Integridad y Coherencia**: Las personas actúan de manera coherente con sus valores y decisiones previas. Los estafadores pueden aprovechar iniciando con solicitudes aparentemente inofensivas que, con el tiempo, los llevan a compromisos más riesgosos.
- **Distracción**: Cuando un usuario está enfocado en algo que capta su atención, es más fácil manipularlo sin que se percate de la amenaza. Los atacantes suelen explotar también este estado para ejecutar sus engaños sin ser detectados.
- **Tiempo:** La presión del tiempo disminuye la capacidad de evaluar situaciones de manera racional. Los estafadores utilizan plazos cortos o escenarios de urgencia para empujar a la víctima a tomar decisiones apresuradas sin considerar los riesgos.

La evidencia empírica respalda la idea de que la forma más efectiva de prevenir ataques de ingeniería social es mediante la educación y capacitación de las posibles víctimas sobre las tácticas de manipulación. Para ello, es fundamental que las personas adquieran el conocimiento necesario sobre las técnicas de ingeniería social para detectar y responder a estos ataques, evitando así ser engañadas y manipuladas." (Syafitri, et al., 2022; Parthy & Rajendran, 2019).

Dado lo anterior, se presentó la siguiente **explicación** sobre las técnicas de ingeniería social presente en los correos electrónicos fraudulentos **(Figura 3):**

Figura 3: Campaña que incluye *Explicación* sobre técnicas de ingeniería social



c) Diseño Enfoque basado en juegos ("Juego")

El diseño del Juego propuesto se basó en el juego Jigsaw: Test de *Phishing*[6], una herramienta interactiva desarrollada por Google, diseñada para evaluar y mejorar la capacidad de los usuarios para identificar correos electrónicos de *phishing*. El test Jigsaw presenta a los participantes una serie de ocho correos electrónicos, algunos legítimos y otros fraudulentos, y les pide que determinen si cada uno es auténtico o un intento de *phishing*. Después de cada respuesta, el test proporciona retroalimentación detallada, señalando las pistas que indican la legitimidad o falsedad del correo, como la dirección del remitente, enlaces sospechosos o errores gramaticales.

En el juego diseñado para este experimento, se utilizaron tres correos electrónicos relacionados con comunicaciones de bancos locales: dos correos fraudulentos—uno de Banco Estado y otro de Banco de Chile, obtenidos del sitio web del CSIRT del Ministerio del Interior y Seguridad Pública—y un correo legítimo de Banco Santander, extraído de una comunicación real de la institución.

La dinámica del juego consistió en la presentación secuencial de los correos, con un nivel de dificultad progresivo:

- **Correo de Banco Estado** (fraudulento, fácil de detectar): Presentado en primer lugar, este correo contenía múltiples elementos sospechosos previamente señalados en la campaña, lo que facilitaba su identificación como fraudulento.
- **Correo de Banco Santander** (legítimo, dificultad media): A continuación, los participantes evaluaban este correo, cuya autenticidad podía generar dudas, ya que, aunque era legítimo, incluía una oferta de duración limitada, un recurso común en correos fraudulentos.
- **Correo de Banco de Chile** (fraudulento, difícil de detectar): Finalmente, los encuestados analizaban este correo, diseñado para ser más difícil de identificar como fraudulento debido a la presencia de pocos elementos sospechosos.

Esta secuencia permitió evaluar la capacidad de los participantes para reconocer patrones de fraude en correos electrónicos de distinta complejidad.

Después de analizar cada correo, los participantes debían evaluar la probabilidad de que el correo exhibido fuera una estafa en línea utilizando una escala de Likert de 5 puntos.

A continuación, debían indicar el número de elementos sospechosos que identificaron en el correo para justificar su respuesta. De manera opcional, podían describir específicamente cuáles eran esos elementos.

A diferencia del juego Jigsaw, este test no proporciona retroalimentación inmediata después de cada respuesta, un elemento que suele ser clave en los juegos serios para fortalecer el aprendizaje. No obstante, se espera que su naturaleza interactiva genere un alto nivel de inmersión en la experiencia, un fenómeno conocido como "transporte", lo que facilita la persuasión. De acuerdo con el Modelo de Probabilidad de Elaboración Extendida, cuando una narrativa persuasiva se presenta a través de una dinámica lúdica—como en este caso, mediante la aplicación de las recomendaciones de la campaña en escenarios prácticos—se incrementa la implicación del individuo. Al estar inmerso en la historia, el usuario tiende a adoptar una postura menos crítica respecto al contenido, lo que reduce la generación de contraargumentos y disminuye la resistencia al mensaje de la campaña (Shrum, 2004; Slater & Rouner, 2002).

[6] https://phishingquiz.withgoogle.com/?hl=es

Figura 4: Correo 1. Banco Estado (fraudulento)

¿Puede detectar si el siguiente correo es una estafa en línea?

Asunto: Aviso Cuenta Temporalmente Suspendida!

B BancoEstado <apache@ippo.com>
Para: Luis Gómez <luis.gomez8000@gmail.com>
para mí

BancoEstado

Estimado Luis,

BancoEstado su clave de internet a vencido Su cuenta se encuentra **SUSPENDIDA** hasta la correcta validacion de sus datos.

realizada la validacion su cuenta sera activada obteniendo los beneficios de banca por internet.

Recuerde que solo tiene 48 horas despues de la fecha de vencimiento para realizar este proceso mediante el enlace que se le proporciona nuestra Banca por internet, de lo contrario su cuenta sera inhabilitada y tendra que acercarse a la sucursal mas cercana para su verificacion respectiva.

Evite el bloqueo desde **aqui**.

Ingrese.Aqui.

Este es un correo electrónico generado automáticamente. Por favor no responder.

Por tu seguridad, sigue estos consejos:

- Nunca compartas tus claves de tarjetas y de acceso a Banca en Línea o Aplicación, ni tus códigos de autorización.
- Siempre ingresa a www.bancoestado.cl, asegurándote que la dirección esté bien escrita.

Juntos contra el fraude

Conoce más **recomendaciones de seguridad de BancoEstado** en www.bancoestado.cl

Síguenos en @bancoestado — Servicio WhatsApp oficial +569 5894 2219

De conformidad al artículo 28 B de la Ley 19.496 sobre Protección de los Derechos de los Consumidores, donde se regula el envío de correo masivos. Si usted no quiere recibir nuevos mensajes desde esta dirección, debe pinchar en el link al final de este correo para no recibir nuevos e-mail. Se deja constancia que los datos de contacto de este envío (direcciones, teléfonos, direcciones electrónicas, etc.) son reales y correctos y su e-mail ha sido extraído a través de medios mecánicos o tecnológicos desde nuestras propias bases de datos, sitios públicos de internet e impresos de publicidad.

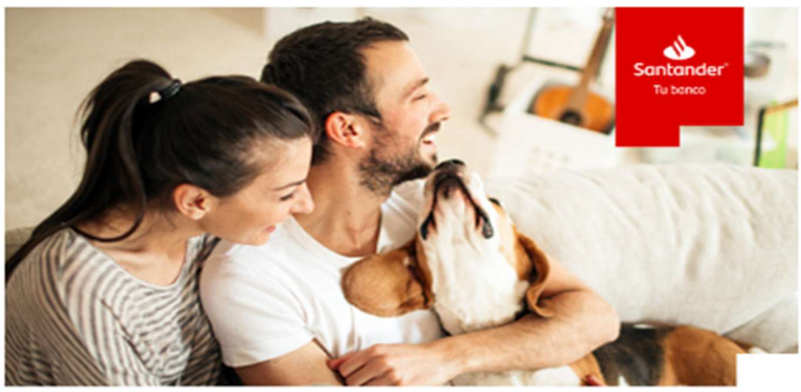
Fuente: CSIRT (8FPH24-00943: *Phishing* que suplanta a Banco Estado)

Figura 5: Correo 2. Banco Santander (legítimo)

¿Puede detectar si el siguiente correo es una estafa en línea?

Asunto: ¡Tu Crédito de Consumo 100% online!

Banco Santander <banco@mensajeria.santander.cl>
Para: Luis Gómez <luis.gomez8000@gmail.com>
para mí



Estimado Luis Alberto Gómez Morel:

Solicita tu Crédito de Consumo 100% online en Santander.cl

Tienes un Crédito preaprobado de hasta:

$18.590.000

15% DCTO.
En Tasa de interés.

*Descuento por monto igual o superior a $5.500.000 líquidos
Oferta exclusiva para contrataciones digitales
Campaña vigente desde el 1 al 30 de abril de 2024.

¡Revisa los descuentos diarios que tenemos para ti!

Solicítalo en **Santander.cl** o en tu **App Santander.**

Fuente: Banco Santander.

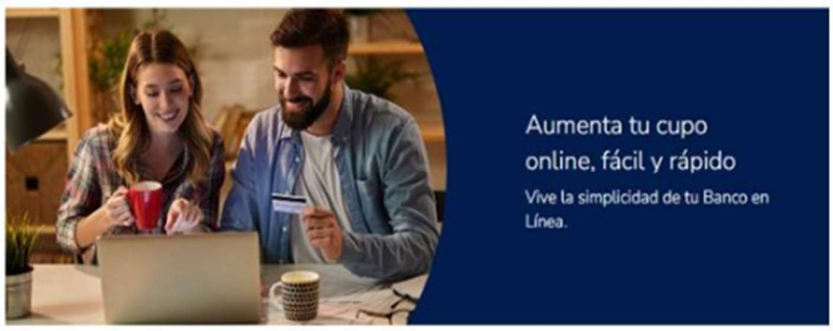
Figura 6: Correo 3. Banco de Chile (fraudulento)

¿Puede detectar si el siguiente correo es una estafa en línea?

Asunto: Consulta si tienes aprobado un aumento de cupo

BC Banco Chile <bancochile@enlinea.cl>
Para: Luis Gómez <luis.gomez8000@gmail.com>
para mí

Si hay problemas con el modo en que se muestra este mensaje, haga clic aquí para verlo en un explorador web.



Consulta si tienes pre aprobado un aumento de cupo

BANCO EN LÍNEA

VIGENCIA

Vigencia desde el 1 al 30 de mayo de 2024.

Vigencia desde el 1 al 30 de mayo de 2024. Otorgamiento de aumento de cupo de Línea de Crédito y/o Tarjeta de Crédito señalado sujeto a que se mantengan condiciones comerciales y financiera del cliente consideradas al momento de la evaluación.

Fuente: CSIRT (8FPH23-00894-01: *Phishing* que suplanta a Banco de Chile).

4.2.4 Condiciones Experimentales

Para evaluar las hipótesis de investigación del estudio, se diseñaron ocho condiciones experimentales, cada una con diferentes combinaciones de los factores analizados **(Tabla 1).** Los participantes fueron asignados de manera aleatoria y estratificada según género y edad a una de las siguientes ocho condiciones experimentales:

- **Tratamiento 1 (Control Experimento 1):** Presenta una campaña estándar, seguida del Test de *Phishing*.
- **Tratamiento 2**: Presenta una campaña reforzada con mensajes motivacionales, los cuales enfatizan la autoeficacia y la gravedad percibida, seguida del Test de *Phishing* (Juego).
- **Tratamiento 3**: Presenta una campaña reforzada con explicaciones sobre técnicas de ingeniería social utilizadas en correos fraudulentos, seguida del Test de *Phishing* (Juego).
- **Tratamiento 4:** Presenta una campaña reforzada con explicaciones sobre técnicas de ingeniería social y mensajes motivacionales enfocados en la autoeficacia y la gravedad percibida, seguida del Test de *Phishing* (Juego).
- **Tratamiento 5 (Control Experimento 2):** Presenta una campaña estándar, sin refuerzo adicional.
- **Tratamiento 6:** Presenta una campaña reforzada con mensajes motivacionales, los cuales enfatizan la autoeficacia y la gravedad percibida.
- **Tratamiento 7:** Presenta una campaña reforzada con explicaciones sobre técnicas de ingeniería social utilizadas en correos fraudulentos.
- **Tratamiento 8:** Presenta una campaña reforzada con explicaciones sobre técnicas de ingeniería social y mensajes motivacionales enfocados en la autoeficacia y la gravedad percibida.

Este diseño experimental permite evaluar los efectos diferenciados de las estrategias de persuasión utilizadas en la campaña, tanto en la capacidad de detección de fraudes (Experimento 1) como en la disposición a adoptar medidas preventivas (Experimento 2).

Tabla 1. Correspondencia entre Factores y Tratamientos

	Factores		
Controles y Tratamientos	**Explicación**	**Motivación**	**Juego**
1 (Control Experimento 1)	0	0	1
2	0	1	1
3	1	0	1
4	1	1	1
5 (Control Experimento 2)	0	0	0
6	0	1	0
7	1	0	0
8	1	1	0

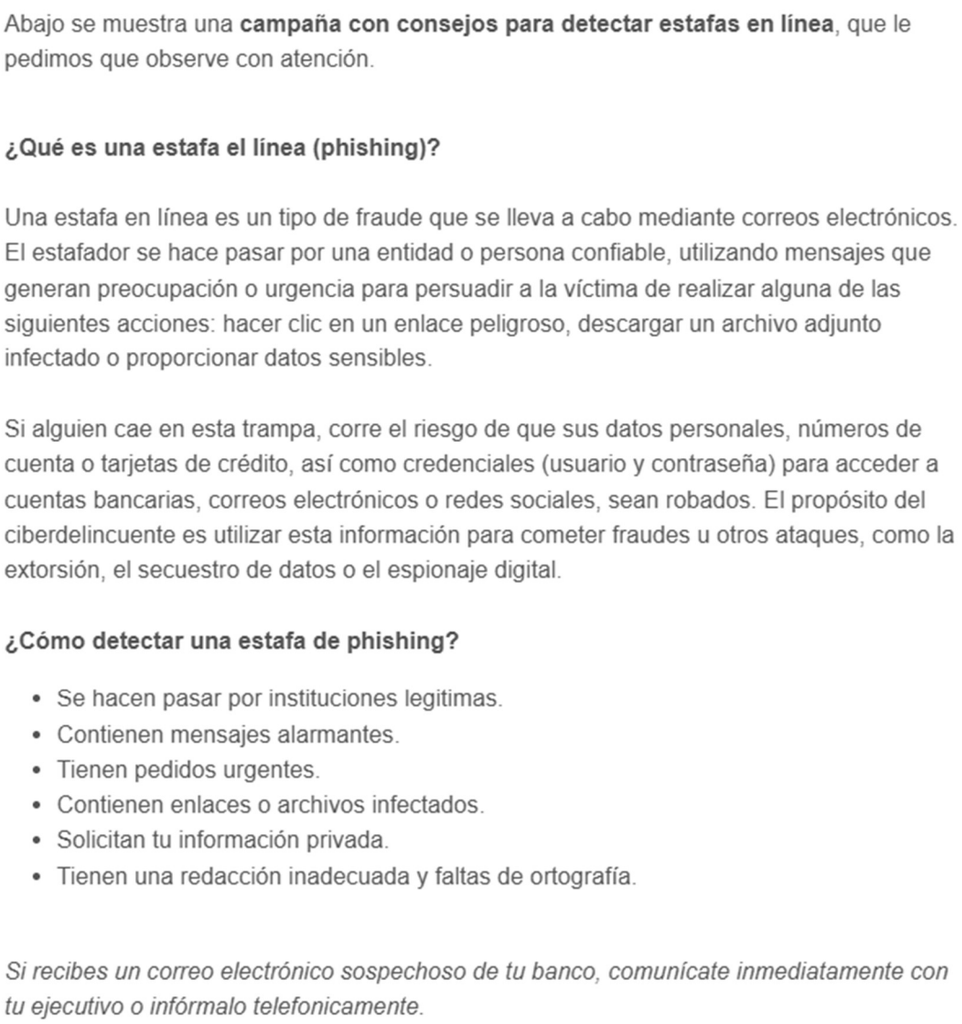
Adicionalmente, se incluyeron dos condiciones experimentales adicionales, diseñadas como líneas base que podían servir como controles (contrafactuales) para los Experimentos 1 y 2, respectivamente:

- **Tratamiento 0 (Control 2, Experimento 1):** Los participantes de este grupo (T0) no fueron expuestos a ningún mensaje previo; en su lugar, pasaban directamente al Test de *Phishing* (Juego).

- **Tratamiento 9 (Control 2, Experimento 2):** Los participantes de este grupo fueron expuestos a una campaña diferente a la campaña estándar presentada en los demás grupos. Además de incluir las seis señales clave para detectar una estafa de *phishing* (presentes en la campaña estándar), esta campaña iniciaba con una descripción del concepto y modus operandi del *phishing* y finalizaba con una recomendación en caso de exposición. El contenido de esta campaña fue elaborado a partir de una comunicación real obtenida del sitio web de un banco chileno. Asimismo, carecía de elementos estéticos atractivos, replicando el diseño de una campaña típica publicada en la página web de un banco local **(Figura 7).** Adicionalmente, los participantes de este grupo no recibieron el test de *phishing* (Juego), ni los otros factores en evaluación (*Explicación* y *Motivación*).

Estas condiciones permitieron evaluar los efectos de las intervenciones experimentales en comparación con escenarios en los que los participantes no recibían información previa o eran expuestos a una campaña más similar a las presentadas en las páginas web de los bancos locales.

Figura 7: Campaña estándar de páginas Web institucionales



Abajo se muestra una **campaña con consejos para detectar estafas en línea**, que le pedimos que observe con atención.

¿Qué es una estafa el línea (phishing)?

Una estafa en línea es un tipo de fraude que se lleva a cabo mediante correos electrónicos. El estafador se hace pasar por una entidad o persona confiable, utilizando mensajes que generan preocupación o urgencia para persuadir a la víctima de realizar alguna de las siguientes acciones: hacer clic en un enlace peligroso, descargar un archivo adjunto infectado o proporcionar datos sensibles.

Si alguien cae en esta trampa, corre el riesgo de que sus datos personales, números de cuenta o tarjetas de crédito, así como credenciales (usuario y contraseña) para acceder a cuentas bancarias, correos electrónicos o redes sociales, sean robados. El propósito del ciberdelincuente es utilizar esta información para cometer fraudes u otros ataques, como la extorsión, el secuestro de datos o el espionaje digital.

¿Cómo detectar una estafa de phishing?

- Se hacen pasar por instituciones legitimas.
- Contienen mensajes alarmantes.
- Tienen pedidos urgentes.
- Contienen enlaces o archivos infectados.
- Solicitan tu información privada.
- Tienen una redacción inadecuada y faltas de ortografía.

Si recibes un correo electrónico sospechoso de tu banco, comunícate inmediatamente con tu ejecutivo o infórmalo telefonicamente.

4.2.5. Variables dependientes

Experimento 1:

Como se mencionó previamente, el Experimento 1 evaluó el impacto de las campañas en la detección de correos fraudulentos y legítimos. En este contexto, los resultados del juego (test de *phishing*) se consideraron variables dependientes para medir la efectividad de la campaña.

Se midieron dos variables dependientes:

1. La **probabilidad de clasificar un correo correctamente** es la respuesta a la pregunta "¿Qué tan probable es que el correo anterior sea una estafa en línea?", que se hizo a los participantes tras la presentación de cada uno de los tres correos electrónicos a clasificar, que podían ser fraudulentos o legítimos.

La variable dependiente "probabilidad" se construyó promediando las respuestas obtenidas para los tres correos presentados. Estas respuestas fueron registradas en una escala de Likert de 5 puntos, que oscilaba entre "muy improbable" y "muy probable". Por otro lado, cuando una persona evalúa un correo legítimo, dado que la pregunta se formula en términos de la probabilidad de que sea una estafa, la respuesta correcta al clasificarlo como legítimo es "muy improbable". Por ello, al codificar las respuestas, se invirtió la escala, de modo que, independientemente del tipo de correo, un valor superior reflejara una mayor precisión en la clasificación.

2. El **número de elementos detectados** es la respuesta a la pregunta "¿Cuántos elementos detectó en el correo para justificar su respuesta?" Si bien basta con identificar un solo elemento sospechoso para clasificar un correo como fraudulento, esta variable permite evaluar el nivel de detalle (minuciosidad) en que incurren los participantes al tomar su decisión de clasificación.

Experimento 2:

El Experimento 2 evaluó el impacto de las campañas en fomentar una mayor disposición de las personas a adoptar preventivamente las recomendaciones de seguridad. Para ello, se midieron las variables actitud hacia las medidas de detección de estafas en línea ("*actitud*") e intención de protegerse en el futuro ("*intención*"), ambas variables consideradas como precursoras del comportamiento observado. Se midieron además variables intermedias, asociadas al Canal de Persuasión que, a la vez, influyen en los cambios de actitud.

3. Para medir la **actitud hacia las medidas de seguridad** se preguntó: "Ahora, después de ver la campaña, ¿Cuál es su actitud hacia las medidas de detección de estafas en línea?". Se midieron tres ítems en una escala de 1 a 7 puntos, correspondientes a las siguientes opciones de respuesta: desfavorable/favorable, mala/buena y negativa/positiva (Nan et al, 2015). Para construir la variable "actitud", se promediaron las respuestas de estos tres ítems y, posteriormente, la variable se re-escaló de un rango de [1,7] a un rango de [1,5].

En un modelo de persuasión, el término **actitud** se refiere a una predisposición psicológica relativamente estable que influye en la evaluación que una persona hace de un objeto, idea, persona o situación en términos favorables o desfavorables. Las actitudes se componen de tres dimensiones: la cognitiva (creencias sobre el objeto de la actitud), la afectiva (emociones y sentimientos asociados) y la conductual (tendencia a actuar de cierta manera

en relación con el objeto). En el modelo de McGuire y en otros enfoques de la persuasión, la actitud es el estado mental del receptor que puede ser modificado por un mensaje persuasivo. Un mensaje persuasivo efectivo busca generar un **cambio de actitud**, que luego podría traducirse en un cambio de comportamiento (McGuire,1964, Petty er al, 2019).

4. La **intención de protegerse** se evaluó mediante la pregunta: "¿Está de acuerdo con la siguiente afirmación? En el futuro, revisaré las señales de estafas en línea en los correos que reciba antes de responderlos." La respuesta se registró en una escala de Likert de 5 puntos, que iba desde "Totalmente en desacuerdo" hasta "Totalmente de acuerdo"(Nan et al., 2015).

De acuerdo con la Teoría de la Acción Razonada (Fishbein y Ajzen, 1975), el concepto de **intención** se refiere a la decisión consciente de una persona de llevar a cabo una conducta específica. Esta intención es el determinante inmediato del comportamiento y está influenciada por la actitud, entre otros factores.

A la vez, de acuerdo con la Teoría de la Motivación de Protección (Rogers, 1975), el concepto de **intención** se refiere a la decisión consciente de una persona de adoptar comportamientos protectores en respuesta a una amenaza percibida.

5. La atención a la campaña se evaluó a través de dos afirmaciones: "La campaña llamó mi atención" y "La campaña capturó mi atención." Las respuestas se registraron en una escala de Likert de 5 puntos, que iba desde "Totalmente en desacuerdo" hasta "Totalmente de acuerdo" (Hagtvedt & Brasel, 2017). Para construir la variable "atención", se promediaron las respuestas de estos dos ítems.

En el modelo de procesamiento de la información propuesto por William McGuire, la **atención** es un proceso clave en la persuasión, en el que el receptor de un mensaje centra su foco en la información presentada, diferenciándola de otros estímulos en el entorno. La atención no solo depende de la presencia del mensaje, sino también de la motivación del receptor y del contexto en el que se recibe. Este paso es crucial ya que, sin la atención adecuada, el mensaje no puede ser procesado, comprendido ni influir en las actitudes del receptor (McGuire, 1985).

6. Para determinar la **comprensión del material presentado** en la campaña, se evaluaron las siguientes afirmaciones: "Los correos de *phishing* frecuentemente contienen errores gramaticales y ortográficos" y "Los estafadores a menudo intentan crear un sentido de urgencia para que las víctimas actúen rápidamente y sin pensar". Las opciones de respuesta eran: verdadero, falso o ni verdadero ni falso. Esta variable no solo mide la comprensión del contenido de la campaña, sino también el recuerdo (recall) de la información proporcionada.

Para su construcción, las respuestas se recodifican en valores binarios (1 o 0), asignando 1 a las respuestas correctas y 0 a las incorrectas. Posteriormente, se calcula el promedio de las dos preguntas recodificadas, obteniendo así el porcentaje de aciertos. La variable resultante varía entre 0 y 1, donde 0 indica que ninguna respuesta es correcta y 1 significa que ambas respuestas son correctas.

En el modelo de procesamiento de la información de William McGuire, la **comprensión** se refiere a la capacidad del receptor para interpretar y procesar cognitivamente el contenido del mensaje persuasivo después de haberle prestado atención. Este paso es fundamental, ya que una vez que el receptor ha focalizado su atención en el mensaje, es necesario que lo entienda para que pueda evaluarlo y determinar si lo acepta o lo rechaza. En resumen, la comprensión en el modelo de McGuire es el proceso mediante el cual el receptor descifra el

contenido del mensaje, facilitando su evaluación y, en consecuencia, el desarrollo de las etapas posteriores del procesamiento de la información.

La **aceptación de la campaña** se midió a través de tres indicadores previamente utilizados por Nan et al. (2015): Efectividad Percibida, Fuerza Argumental Percibida y Gusto General por la Campaña.

En el modelo de procesamiento de la información de William McGuire, la **aceptación** se refiere al proceso mediante el cual el receptor, después de haber atendido y comprendido un mensaje persuasivo, evalúa y decide si adopta o no la posición o actitud propuesta en dicho mensaje. Este paso es fundamental en la cadena de procesamiento persuasivo, ya que implica una evaluación crítica del contenido y determina si se producirá un cambio de actitud en el receptor (McGuire, 1969, 1985).

7. **Efectividad percibida**. Se evaluó mediante una escala de tres ítems: "¿Qué tan efectiva es la campaña para evitar que las personas sean víctimas de estafas en línea?", "¿Qué tan efectiva es la campaña para motivar a las personas a protegerse de las estafas en línea?" y "¿En qué medida la campaña hace que sienta ganas de protegerse de las estafas en línea?". Las respuestas se registraron en una escala de Likert de 5 puntos, que va de 1 (Nada) a 5 (Mucho).

8. **Fuerza argumental percibida**. Se evaluó mediante una escala de nueve ítems, propuesta originalmente por Zhao et al. (2011), la cual mide la percepción de los participantes sobre distintos aspectos del contenido de la campaña, como su, credibilidad, importancia, interés y solidez, entre otros. Los participantes respondieron las siguientes afirmaciones: "La campaña proporciona recomendaciones creíbles para detectar estafas en línea"; "La campaña entrega recomendaciones convincentes para detectar estafas en líneas"; "La campaña entrega razones importantes para mi"; "La campaña me ayudó a sentirme seguro sobre cómo protegerme de las estafas en línea"; "La campaña ayudaría a mis amigos a detectar estafas en línea; "La campaña me generó interés en protegerme de las estafas en línea". "La campaña me generó interés en no exponerme a las estafas en línea"; "En general, ¿Qué tan de acuerdo o en desacuerdo está con las recomendaciones de la campaña?". Para estos ítems, se utilizó una escala de Likert de 5 puntos, que va de 1 (Totalmente en desacuerdo) a 5 (Totalmente de acuerdo). Además, la última pregunta, "¿La campaña entrega razones sólidas o débiles para prevenir las estafas en línea?", se midió con una escala de 7 puntos, cuyos extremos corresponden a Débiles y Sólidas.

9. **Gusto general por la campaña**. Se preguntó a los participantes "¿Le gustó la campaña que acaba de ver?". Las respuestas se registraron en una escala de Likert de 5 puntos, que va de 1 (Nada) a 5 (Mucho).

Para medir la motivación de protección, se utilizaron dos constructos basados en la Teoría de Motivación de Protección (TMP): gravedad percibida y autoeficacia. La escala para evaluar estos constructos fue desarrollada en un estudio previo de SERNAC (2025b). Cada constructo se midió a través de tres ítems, cuyas respuestas, basadas en el grado de acuerdo con cada afirmación, se registraron en una escala de Likert de 5 puntos (de 1: Totalmente en desacuerdo a 5: Totalmente de acuerdo).

10. **Gravedad percibida.** Este constructo forma parte del proceso de evaluación de la amenaza dentro de la TMP. Se define como la percepción de la magnitud del impacto negativo que podría tener un fraude en línea. Los participantes evaluaron las siguientes afirmaciones: "La pérdida financiera resultante de una estafa en línea sería un problema

grave para mí."; "Una estafa en línea podría dañar gravemente mi situación financiera."; "Que alguien ataque mis sitios financieros sería muy perjudicial para mí.".

11. **Autoeficacia percibida.** Este constructo pertenece al proceso de evaluación del afrontamiento dentro de la TMP. Se define como la percepción de la propia capacidad para implementar las medidas de protección recomendadas. Los participantes evaluaron las siguientes afirmaciones: "La campaña me hizo sentir que cuento con los conocimientos necesarios para protegerme de las estafas en línea"; "La campaña me hizo sentir que cuento con los conocimientos necesarios para protegerme de las estafas en línea"; "La campaña me hizo sentir capaz de seguir las recomendaciones incluso sin nadie cerca que me muestre como".

4.2.6 Variables de Control

A continuación, se detallan las variables empleadas como controles en los distintos análisis del estudio. En primer lugar, al inicio de la encuesta se registraron el género y la edad de los participantes para llevar a cabo una aleatorización estratificada y, posteriormente, dichas variables se utilizaron en análisis de heterogeneidad. Además, todas las variables de control se emplearon en pruebas de balance y en un análisis de robustez, con el fin de evaluar la estabilidad de los resultados y controlar posibles sesgos, garantizando así la comparabilidad estadística entre los grupos experimentales.

Las variables de control utilizadas son:

Víctima de fraude en línea ("Víctima"). Se preguntó a los participantes: "¿Ha sido víctima de estafas en línea durante el último año?", con opciones de respuesta Sí o No. Esta variable es relevante, ya que un estudio previo (SERNAC, 2025b) encontró una correlación negativa entre la intención de protegerse y la probabilidad de haber sufrido fraudes: quienes menos se protegen han sido víctimas de estafas en el pasado.

Experiencia en el uso de internet ("Experiencia"). Se preguntó a los participantes: "¿Se considera experimentado(a) en el uso de internet? Considere su experiencia en el uso de email, sitios web, banca en línea, redes sociales, etc." Esta variable mostró una alta correlación con la motivación de protección en SERNAC (2025b). Las respuestas se registraron en una escala de Likert de 5 puntos, que va de 1 (Muy por debajo del promedio) a 5 (Muy por encima del promedio).

Necesidad de cognición ("Cognición"). Esta variable se midió a través de una escala de seis afirmaciones (ítems), desarrollada por Coelho et al. (2020), las cuales se registraron en una escala de Likert de 5 puntos (de 1: Totalmente en desacuerdo a 5: Totalmente de acuerdo): a) Prefiero problemas complejos a problemas simples; b) Me gusta tener la responsabilidad de manejar una situación que requiere mucho pensamiento; c) Disfruto mucho una tarea que implique idear nuevas soluciones a problemas; d) Prefiero una tarea que sea intelectual, difícil e importante a una que sea algo importante pero que no requiera mucho pensamiento; e) Pensar no es mi idea de diversión; f) Prefiero hacer algo que requiera poco pensamiento que algo que seguramente desafíe mis habilidades de pensamiento.

La necesidad de cognición de es un rasgo de personalidad estable que refleja la tendencia de las personas a disfrutar y participar en actividades cognitivas que requieren esfuerzo. (Cacioppo y Petty, 1982). Quienes tienen una alta necesidad de cognición, buscan información, reflexionan sobre ella y responden de manera más sustantiva a situaciones que exigen pensamiento y resolución de problemas. En cambio, quienes tienen una baja necesidad de cognición suelen recurrir a atajos mentales o heurísticas para interpretar el

mundo. Por lo tanto, se espera que las personas con una mayor necesidad de cognición tengan actitudes más positivas hacia las situaciones que requieren razonamiento y resolución de problemas, y que respondan de forma más sustantiva a dichas situaciones (Cacioppo, et al, 1996). Dado lo anterior, se podría esperar que, en el marco de este estudio, los tratamientos más complejos (aquellos que combinan todos los factores: Explicación, motivación y juego) tengan un efecto mayor en las personas con una mayor necesidad de cognición.

Porcentaje de encuestados que respondió la encuesta en un teléfono móvil ("Móvil"): La variable "Móvil" es un paradato recopilado por Qualtrics, que indica si un participante respondió la encuesta desde un dispositivo móvil o una computadora. Algunos estudios han reportado diferencias en el comportamiento de los encuestados según el dispositivo utilizado, como tiempos de respuesta más largos y mayores tasas de abandono en móviles, posiblemente debido a distracciones más frecuentes (Liebe et al., 2015; Herzing, 2019; Mavletova & Couper, 2015).

Además, responder desde un celular puede afectar la calidad de los datos debido a diversas limitaciones tecnológicas y cognitivas. La interfaz táctil incrementa la posibilidad de errores en la selección de respuestas (Herzing, 2019) y puede inducir un sesgo de orden de respuesta, ya que los participantes tienden a seleccionar las primeras opciones debido al desplazamiento vertical en pantallas móviles (Tourangeau et al., 2018). Además, se ha observado una mayor incidencia de "straightlining" en móviles en comparación con computadoras, donde los encuestados eligen sistemáticamente las mismas opciones dentro de una columna o conjunto de preguntas similares (Struminskaya et al.,2015).

Adicionalmente, en el estudio se consideraron diversas variables socioeconómicas para caracterizar a los participantes y analizar posibles diferencias en los efectos del tratamiento. Como se mencionó previamente, se incluyó el **género**, preguntando a los participantes con cuál se identificaban (femenino, masculino, otro o prefiero no decirlo). La **edad** se registró en tramos: 18-29 años, 30-44 años, 45-59 años y 60 años o más. Asimismo, se recopiló información sobre los **ingresos mensuales**, considerando todas las fuentes de ingresos y clasificándolos en seis categorías, desde menos de $250.000 hasta más de $2.000.000. Finalmente, se indagó sobre el **nivel educacional alcanzado**, diferenciando entre educación básica incompleta, básica completa, media completa, educación técnica superior, universitaria y postgrado.

4.3 Implementación

4.3.1 Screening

El estudio se llevó a cabo mediante un experimento en línea, utilizando un panel compuesto por consumidores pertenecientes a la base de datos de reclamos del SERNAC. A partir de esta base, se extrajo una muestra aleatoria mediante un muestreo aleatorio estratificado por género y edad, siguiendo las proporciones poblacionales de cada estrato.

Los participantes seleccionados fueron contactados por correo electrónico, donde recibieron una invitación con un enlace a un cuestionario en línea implementado a través de la plataforma *Qualtrics.* Esta plataforma facilitó tanto la presentación del experimento como la recolección de datos y la aleatorización de los participantes.

El proceso incluyó dos fases piloto:
Prueba piloto inicial: Se realizó entre el 5 y 7 de noviembre de 2024 con 48 estudiantes universitarios (n=48), lo que permitió mejorar el cuestionario y verificar el proceso de recolección de datos.
Piloto de la versión final: Se llevó a cabo entre el 21 y 27 de noviembre con 507 participantes (n=507). Su propósito fue determinar los tamaños muestrales óptimos para la evaluación, mediante un análisis de poder estadístico.

Tras estas pruebas, el levantamiento final de datos se realizó entre el 28 de noviembre y el 16 de diciembre de 2024. En total, se recopilaron 6.044 respuestas válidas, es decir, encuestas completamente finalizadas. Esta cifra incluye los datos de dos grupos experimentales adicionales (T0 y T9), los cuales no fueron considerados en los análisis principales. Excluyendo estos grupos, la muestra final utilizada ascendió a 4.703 participantes.

En términos de participación, la tasa de respuesta fue del 3%[7]. Además, el 36% de quienes iniciaron la encuesta la completaron[8]. Entre las encuestas terminadas, el 60% de los participantes utilizó un teléfono móvil para contestarlas, y la duración mediana del cuestionario fue de 11 minutos.

4.3.2 Flujo de la Encuesta En Línea

A continuación, se describe el flujo de la encuesta en línea **(Figura 8).**
La encuesta comenzaba con una sección de **consentimiento informado**, en la que se explicaba a los participantes el objetivo del estudio, la naturaleza voluntaria de su participación y la confidencialidad y anonimato de la información recopilada.

Luego, los participantes debían indicar su género y tramo de edad. Con base en estas respuestas, eran asignados aleatoriamente a una de las diez condiciones experimentales, que incluían los ocho grupos del diseño factorial y dos grupos de control adicionales. La asignación se realizó mediante un **procedimiento aleatorio estratificado** por género y

[7] Porcentaje de las personas contactadas que respondieron completamente la encuesta.

[8] En las pruebas de robustez de los resultados se repitieron las estimaciones usando datos de las encuestas incompletas, para verificar que este *desgaste* (*attrition*) no afecte los resultados. Se dice que hay *desgaste* en un experimento cuando no se puede recolectar datos de resultados de individuos que formaban parte de la muestra original. En este caso, cuando un individuo abandona la encuesta, sólo se tienen datos de las respuestas iniciales, anteriores al abandono.

edad, asegurando una distribución equitativa de los participantes entre las condiciones experimentales.

Exposición a los Tratamientos. Una vez asignados a su grupo experimental, los participantes fueron expuestos a los tratamientos. En primer lugar, se les presentó una lista de seis señales características de correos fraudulentos, diseñadas para ayudar en su detección. Dependiendo de la condición experimental asignada, estas recomendaciones podían estar reforzadas con:

- ✓ Una **Explicación**, que consistía en un texto adicional detallando brevemente las técnicas de ingeniería social utilizadas en los correos
- ✓ Un **mensaje motivacional** para incentivar la adopción de las recomendaciones.
- ✓ Un **juego interactivo**, que implicaba una prueba de clasificación de correos fraudulentos.

Así, los participantes en los grupos T1, T2, …, T8 recibieron un mensaje con recomendaciones para detectar correos fraudulentos, el cual podía incluir una Explicación y/o un mensaje motivacional. En tanto, la mitad de los grupos experimentales (T1 a T4) participó en el juego interactivo, mientras que la otra mitad (T5 a T8) no participó.

En el caso de T0 (Control 2, Experimento 1), los participantes de este grupo no fueron expuestos a ningún mensaje previo; en su lugar, pasaban directamente al Test de *Phishing* (Juego). En caso de T9 (Control 2, Experimento 2), Los participantes de este grupo fueron expuestos a una campaña diferente a la campaña estándar presentada en los demás grupos. La nueva campaña replicaba, en parte, el diseño y contenido de una campaña típica publicada en la página web de un banco local.

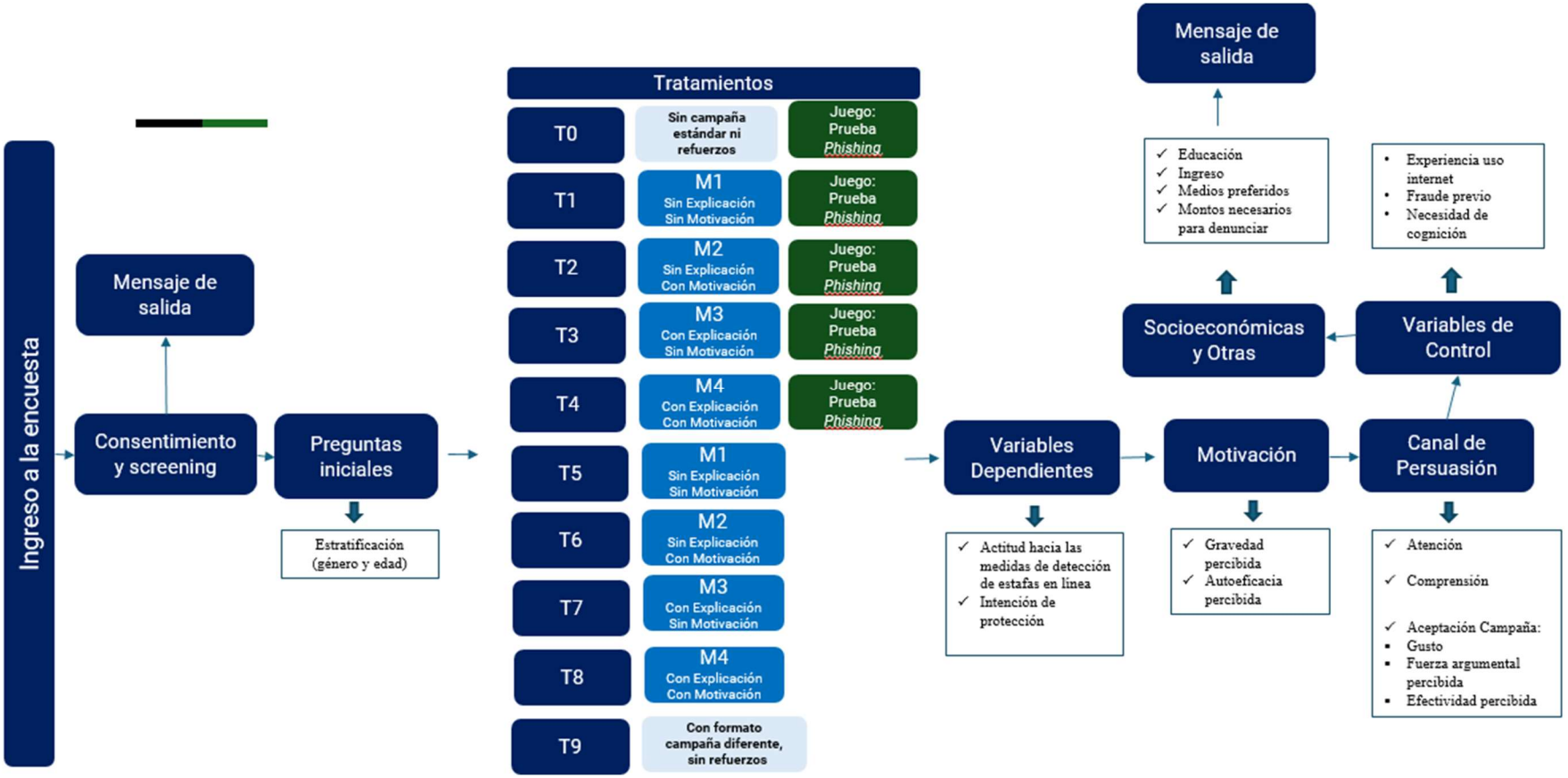
Medición de Resultados. Tras la exposición a los tratamientos, los participantes respondieron preguntas relacionadas con los resultados finales (*Actitud* hacia las medidas de seguridad, *Intención* de protegerse) e intermedios asociados al Modelo de Comunicación y Persuasión (*Atención* prestada al contenido, *Comprensión* del material presentado, Aceptación de la Campaña (*Efectividad percibida, Fuerza argumental percibida, Gusto general* por la campaña) y Conductores de la Motivación (*Gravedad percibida* del fraude*, Autoeficacia percibida*) [9].

Variables de Control y Preferencias Posteriormente, los participantes respondieron preguntas relacionadas con variables de control, tales como: i) Experiencia con el uso de internet; ii) Historial de victimización por fraude en el último año; iii) Necesidad de cognición; iv) Caracterización socioeconómica (nivel de educación y tramo de ingreso).
Asimismo, se indagó sobre las preferencias de los participantes en relación con las campañas informativas, incluyendo los Medios a través de los cuales preferirían recibir información sobre fraudes y los Montos mínimos defraudados que los motivarían a denunciar ante su institución financiera y/o PDI.

Cierre de la Encuesta. Finalmente, tras completar la encuesta, se agradeció la participación de los encuestados. Cuando correspondía, se les informó que los correos fraudulentos presentados en la encuesta fueron obtenidos aleatoriamente del sitio web del CSIRT, una agencia gubernamental dependiente del Ministerio del Interior y Seguridad Pública, y que todas las instituciones financieras están expuestas al fenómeno de la suplantación de identidad.

[9] El grupo T0, dado que no fue expuesto a ninguna campaña, no respondió las preguntas de evaluación de la campaña, sino que pasó directamente a las preguntas de caracterización socioeconómica.

Figura 8: Flujo de la Encuesta en Línea



5. Resultados Principales

5.1 Análisis de la muestra

5.1.1 Descripción de la muestra

A continuación, se describen las características generales de los participantes considerados en este estudio. En cuanto a su caracterización socioeconómica, el 48% de los encuestados se identificó con el género femenino, mientras que el 51% con el género masculino. El 1% restante se identificó con otro género o prefirió no responder **(Gráfico 1).** Respecto al rango etario, aproximadamente la mitad de los encuestados declaró tener entre 18 y 44 años, mientras que el otro 50% tenía 45 años o más. Los rangos de edad más frecuentes fueron de 30 a 44 años (42%) y de 45 a 59 años (32%) **(Gráfico 2).** En cuanto al nivel educacional, el 85% indicó haber cursado educación superior técnica, universitaria o de postgrado, mientras que solo el 14% tenía educación básica o media **(Gráfico 3).** Respecto al nivel de ingresos, el 61% de los encuestados reportó un ingreso mensual de $901.000 o más **(Gráfico 4).**

Gráfico 1. Encuestados según género (%)

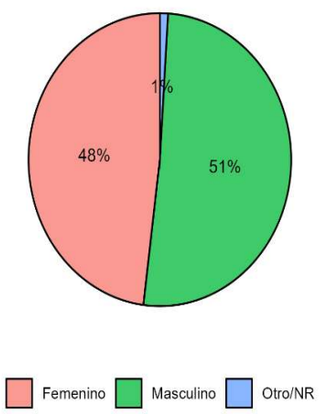
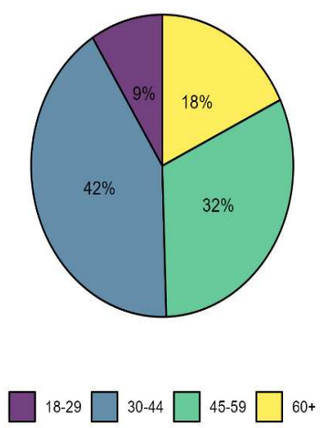


Gráfico 2. Encuestados según edad (%)



Fuente: Elaboración propia.

Gráfico 3. Encuestados según Nivel Educacional (%)

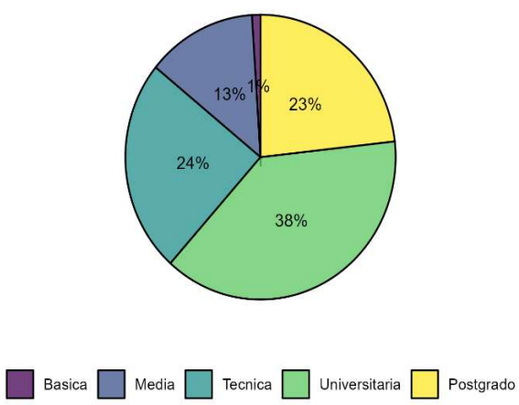
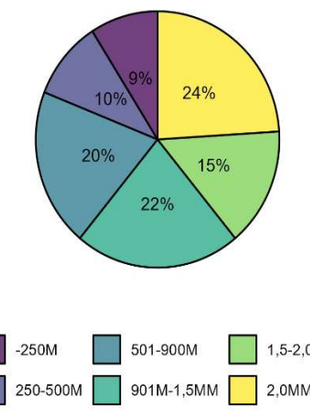


Gráfico 4. Encuestados según Nivel de ingresos (%)



Fuente: Elaboración propia

Se preguntó a los participantes: "¿A través de qué medio preferiría acceder a las campañas de seguridad provenientes de organismos públicos e instituciones financieras?", permitiéndoles elegir entre una lista de medios de comunicación. El siguiente gráfico presenta los medios más seleccionados, destacándose el email, el video, el sitio web del SERNAC y el sitio web de las instituciones financieras **(Gráfico 5).**

Cabe señalar que, al analizar las respuestas según género, se observan las mismas cuatro primeras preferencias en el mismo orden. Un patrón similar se encuentra al segmentar por grupo etario, aunque con variaciones en el orden de preferencia. Por ejemplo, el grupo de 18-29 años prioriza el video, mientras que el grupo de 60 años o más prefiere el email como principal medio de acceso a las campañas.

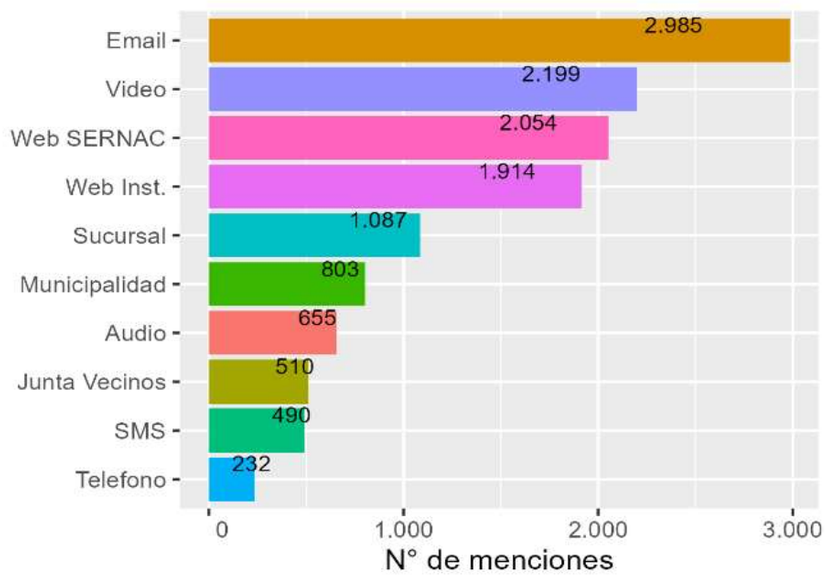
Se consultó a los encuestados: "¿A partir de qué monto defraudado decidiría usted realizar la denuncia en su banco?", considerando dos escenarios:

- Si la denuncia puede realizarse directamente en el sitio web del banco o por teléfono, presentando una declaración jurada (monto 1).
- Si, además, es necesario presentar la denuncia de forma presencial ante Carabineros o la Policía de Investigaciones (PDI) (monto 2).

Esta pregunta surge debido a que, antes de la última modificación a la ley de fraude en mayo de 2024 (Ley N° 21.673), los consumidores se encontraban en el primer escenario. Sin embargo, con la nueva normativa, deben enfrentar el segundo, que implica un proceso más engorroso. Por lo tanto, el objetivo de la pregunta es valorar el costo de oportunidad asociado a cada situación.

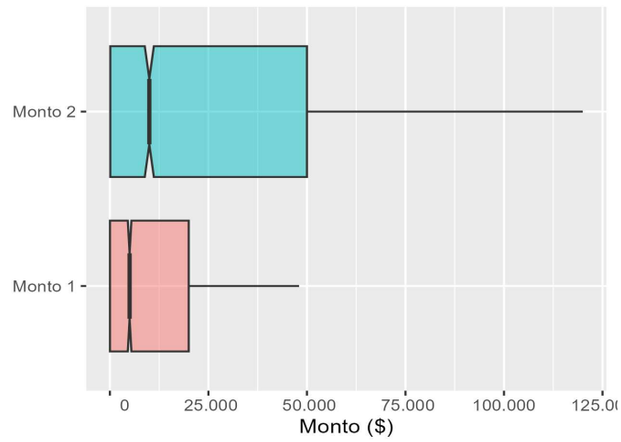
Los resultados muestran que la mediana del monto señalado para el primer escenario fue de $5.000, con un rango entre $20 y $20.000 (primer y tercer cuartil). En el segundo escenario, la mediana del monto fue de $10.000, con valores que oscilan entre $100 y $50.000. En general, se concluye que no es necesario un monto defraudado elevado para que las víctimas decidan denunciar un fraude **(Gráfico 6).**

Gráfico 5. Medios de comunicación preferidos



Fuente: Elaboración propia.

Gráfico 6. Montos necesarios para denunciar ($)



Fuente: elaboración propia.

5.1.2 Prueba de Balance

Para garantizar la validez de la comparación entre grupos, es fundamental que estos sean estadísticamente equivalentes en características clave antes de la intervención. Para ello, se realizó una **prueba de balance**, cuyo propósito es confirmar que cualquier diferencia en los resultados pueda atribuirse exclusivamente al tratamiento aplicado. En particular, esta prueba evalúa si la asignación de los participantes fue realmente aleatoria y si existen diferencias sistemáticas entre los grupos que pudieran influir en los resultados.

Para llevar a cabo la prueba de balance, se analizaron variables socioeconómicas (género, edad, nivel educacional y nivel de ingreso) y variables de control (*víctima, experiencia, cognición* y *móvil*), las cuales se describen en la sección 4.2.6. La **Tabla 2** presenta los promedios de cada variable para cada uno de los 10 grupos experimentales. En la mayoría de los casos, las cifras se expresan en porcentajes. Por ejemplo, el porcentaje de mujeres en cada grupo de tratamiento oscila entre el 44 % y el 52 %. Las únicas excepciones son las variables de *experiencia* y *cognición*, las cuales se miden en una escala de 1 a 5. Además, la última fila de la tabla muestra el tamaño muestral de cada grupo experimental. El grupo T2 tuvo el menor número de participantes, con un mínimo de 538, mientras que el grupo T0 registró el mayor número, con un máximo de 782 participantes.

Tabla 2. Estadística descriptiva por grupo experimental

Variables de Control	T0: Control 2, Exp. 1	T1: Control 1, Exp.1	T2	T3	T4	T5: Control 1, Exp.2	T6	T7	T8	T9: Control 2, Exp. 2
Mujeres	0,49	0,45	0,44	0,49	0,48	0,50	0,52	0,48	0,50	0,49
Hombres	0,50	0,54	0,55	0,50	0,52	0,50	0,47	0,52	0,50	0,50
Edad: 18-29	0,11	0,09	0,10	0,09	0,09	0,10	0,07	0,08	0,08	0,08
Edad: 30-44	0,44	0,43	0,43	0,43	0,41	0,41	0,41	0,39	0,42	0,37
Edad: 45-59	0,30	0,30	0,30	0,30	0,33	0,31	0,34	0,34	0,31	0,37
Edad: 60 o más	0,16	0,19	0,17	0,18	0,17	0,18	0,18	0,20	0,19	0,18
Ed. básica	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,01	0,01
Ed. media	0,14	0,13	0,10	0,12	0,14	0,13	0,13	0,14	0,12	0,15
Ed. técnica	0,23	0,23	0,25	0,23	0,25	0,23	0,25	0,24	0,24	0,23

Ed. universitaria	0,37	0,40	0,38	0,37	0,38	0,42	0,38	0,36	0,39	0,38
Ed. postgrado	0,24	0,22	0,26	0,27	0,23	0,21	0,22	0,24	0,23	0,23
Ingreso: 250M	0,08	0,10	0,09	0,10	0,11	0,09	0,09	0,08	0,09	0,09
Ingreso: 250M-500M	0,09	0,10	0,08	0,10	0,09	0,10	0,12	0,11	0,09	0,12
Ingreso: 501M-900M	0,18	0,18	0,18	0,18	0,21	0,22	0,21	0,20	0,19	0,18
Ingreso: 900M-1,5MM	0,25	0,23	0,21	0,21	0,20	0,22	0,21	0,24	0,23	0,23
Ingreso: 1,5MM-2,0MM	0,15	0,14	0,18	0,17	0,16	0,14	0,14	0,14	0,15	0,14
Ingreso: 2,0MM o más	0,25	0,24	0,25	0,23	0,23	0,23	0,23	0,23	0,23	0,23
Víctima	0,24	0,24	0,23	0,26	0,25	0,28	0,29	0,25	0,29	0,30
Experiencia	3,50	3,63	3,69	3,72	3,65	3,71	3,60	3,67	3,65	3,60
Cognición	3,80	3,85	3,83	3,86	3,81	3,77	3,78	3,76	3,78	3,70
Móvil	0,57	0.58	0.56	0.56	0.62	0.59	0.63	0.61	0.64	0,65
n	**782**	**561**	**538**	**544**	**542**	**605**	**664**	**627**	**622**	**559**

Aunque el estudio incluyó 10 grupos experimentales, los Experimentos 1 y 2 utilizaron conjuntos de grupos distintos. En el Experimento 1 se emplearon los grupos T1 a T4, junto con T0 como segundo control, mientras que en el Experimento 2 se incluyeron los grupos T1 a T8, con T9 como segundo control. Debido a esta diferencia, las pruebas de balance se realizaron considerando esta distinción **(Tabla 3).** Es importante señalar que los segundos controles no pueden incorporarse en el análisis factorial del ANOVA. Por esta razón, su uso se limita a los modelos basados en tratamiento, donde cumplen dos propósitos: primero, evaluar si una campaña estándar es efectiva o si, por el contrario, no presenta diferencias significativas en comparación con un escenario sin información previa (Experimento 1); y segundo, determinar si los formatos de las campañas (diagramas versus textos) influyen en los resultados.

Las pruebas de balance se realizaron mediante pruebas de diferencia de medias entre los grupos correspondientes a cada experimento, utilizando pruebas de Chi-cuadrado para variables categóricas y pruebas F de ANOVA para variables que se consideraron como continuas (*experiencia y cognición*).

La **tabla 3** muestra los valores p de las pruebas estadísticas utilizadas para evaluar diferencias entre grupos. Como se mencionó previamente, para variables categóricas, se emplea el test de Chi-cuadrado, que permite determinar si existe una asociación significativa entre la variable de control y los grupos. Esto se hace comparando la distribución observada con la esperada bajo la hipótesis nula (Ho: No existe asociación entre las variables analizadas, es decir, la distribución observada de una variable es independiente de la otra).

Para variables continuas, se utiliza el test F de ANOVA, que evalúa si al menos una de las medias de los grupos es significativamente diferente. En este caso, la hipótesis nula establece que no existen diferencias estadísticamente significativas entre las medias de los grupos, por lo que cualquier variación observada sería producto del azar. Si se rechaza la hipótesis nula, se requieren pruebas post hoc para identificar cuáles grupos presentan diferencias significativas.

Si el *p-valor* es alto (por ejemplo, superior a 0.01, 0.05 o 0.10, lo que corresponde a niveles de significancia del 1%, 5% y 10%, respectivamente), no hay suficiente evidencia para rechazar la hipótesis nula. Por el contrario, si el *p-valor* es menor que estos umbrales, se rechaza la hipótesis nula, lo que sugiere que al menos un grupo presenta diferencias significativas o existe una asociación entre las variables analizadas.

En el contexto de una prueba de balance, un p-valor bajo sugiere un posible desbalance en esa variable, lo que podría afectar la validez de los resultados experimentales.

De acuerdo con los resultados de balance presentados en la **Tabla 3**, en los distintos subgrupos (Muestra Total, Experimento 1 y Experimento 2), la mayoría de las variables, como edad, nivel educativo e ingreso, no muestran diferencias significativas, lo que indica una comparabilidad adecuada entre los grupos. No obstante, se identifican desequilibrios en "*Móvil*" (significativo en Experimentos 1 y 2), "*Experiencia*" (en Experimento 1), "*Cognición*" (en Experimento 2) y "*Víctima*" (leve en Experimento 2), lo que sugiere que estos factores podrían afectar los resultados. Dado que un balance adecuado es esencial para la validez del experimento, estas diferencias se abordaron en las pruebas de robustez (sección 5.2.3 del documento) mediante ajustes estadísticos y la inclusión de covariables, con el fin de minimizar sesgos y fortalecer la validez interna del estudio.

Tabla 3. Prueba de Balance

Variables de Control	Muestra Total	Experimento 1		Experimento 2	
	De T0 a T9	De T1 a T4	De T0 a T4	De T1 a T8	De T1 a T9
Mujeres	0.209	0.369	0.346	0,1068	0.150
Hombres	0.225	0.377	0.308	0,1233	0.168
Edad: 18-29	0.439	0.929	0.824	0,5935	0.683
Edad: 30-44	0.291	0.906	0.869	0,8081	0.436
Edad: 45-59	0.131	0.634	0.729	0,5638	0.175
Edad: 60 o más	0.727	0.856	0.680	0,8790	0.930
Ed. básica	0.724	0.785	0.500	0,9313	0.934
Ed. media	0.333	0.179	0.146	0,4525	0.317
Ed. técnica	0.994	0.881	0.940	0,9916	0.989
Ed. universitaria	0.623	0.611	0.737	0,4474	0.549
Ed. postgrado	0.367	0.169	0.283	0,2209	0.303
Ingreso: 250M	0.802	0.697	0.517	0,7115	0.802
Ingreso: 250M-500M	0.266	0.778	0.830	0,5481	0.339
Ingreso: 501M-900M	0.565	0.610	0.732	0,5500	0.577
Ingreso: 900M-1,5MM	0.346	0.516	0.112	0,5984	0.660
Ingreso: 1,5MM-2,0MM	0.366	0.286	0.246	0,2483	0.295
Ingreso: 2,0MM o más	0.995	0.922	0.943	0,9984	0.998
Móvil	0.004 ***	0.117	0.202	0,0233 **	0.006 ***
Experiencia	0.000 ***	0.295	0.000 ***	0,2515	0.129
Víctima	0.051 *	0.640	0.793	0,1101	0.055 *
Cognición	0.042 **	0.653	0.101	0,1353	0.041 **
n	6044	**2185**	2967	**4703**	5262

*Nota: *p<0,1; **p<0,05; ***p<0,01*

5.1.3 Análisis de Poder

El **análisis de poder estadístico** permite evaluar la capacidad de un experimento para detectar efectos reales en los datos, minimizando la probabilidad de cometer errores Tipo II (falsos negativos). Es decir, cuando un experimento no arroja resultados estadísticamente significativos, existen dos explicaciones posibles: a) La hipótesis nula (H0) es verdadera, lo que significa que no hay diferencias o asociaciones significativas entre los grupos experimentales en la variable de interés; b) La hipótesis alternativa (H1) es verdadera, pero el experimento no tuvo suficiente poder estadístico para detectarla (Harrison et al., 2020).

El análisis de poder estadístico considera varios factores clave, como el tamaño del efecto, el nivel de significancia (α) y el poder estadístico ($1-\beta$) para determinar el tamaño de muestra necesario que garantice una alta probabilidad de obtener un resultado significativo si el efecto realmente existe. Un estándar comúnmente aceptado es diseñar experimentos con un poder estadístico de al menos 0,80. Esto significa que, si hay un efecto real con la magnitud especificada, el experimento tiene un 80% de probabilidad de detectarlo y rechazar la hipótesis nula (Harrison et al., 2020).

En este estudio, se establece un nivel de significancia de 0,05 y un poder estadístico de 0,80. Además, con una muestra de n=538 (correspondiente al grupo experimental más pequeño), el experimento tiene la capacidad de detectar diferencias de 0,17 (Cohen's d) en una prueba t de Student, utilizada para evaluar diferencias de medias entre tratamientos. Asimismo, este tamaño de muestra permite detectar diferencias de 0,058 (Cohen's f), lo que equivale a un *f2* de 0,0033 y un *eta2* de 0,0033 en una prueba ANOVA, utilizada para evaluar los efectos de los factores.

Por lo tanto, se puede decir que el experimento tiene una alta capacidad para detectar efectos pequeños, lo que indica un alto poder estadístico. En consecuencia, se puede tener una alta confianza en las inferencias realizadas, tanto al rechazar como al no rechazar las hipótesis nulas.

5.2 Resultados de los Ensayos controlados Aleatorizados

5.2.1 Resultados del Modelo basado en Tratamientos

Experimento 1

El Experimento 1 analiza cómo la inclusión de explicaciones sobre las técnicas de ingeniería social utilizadas en correos fraudulentos ("*Explicaciones*"), dentro de una campaña estándar contra el *phishing*, así como la incorporación de mensajes motivacionales que refuerzan la autoeficacia y la percepción de gravedad ("*Motivación*"), influyen en la probabilidad de identificar un correo fraudulento ("*Probabilidad*") y en *el* número de elementos reconocidos como señales de *phishing*. ("*N° de elementos*")

Para evaluar el experimento, se emplearon dos grupos de control. En primer lugar, se utilizó el grupo experimental T1 como referencia, conformado por participantes expuestos a la campaña estándar, sin refuerzos, quienes posteriormente realizaron el Test de *Phishing*. En segundo lugar, se consideró el grupo experimental T0, cuyos participantes no recibieron ningún mensaje previo y accedieron directamente al test[10].

El desempeño del grupo T0 en la clasificación de correos se interpreta como una línea de base, reflejando el nivel de conocimiento previo de los participantes sobre detección de correos fraudulentos, dado que no recibieron ninguna recomendación durante el experimento. Incluir este grupo como control permite, además, evaluar el impacto de T1, es decir, el efecto de una campaña estándar en la detección de correos fraudulentos.

En la Tabla 4 se presentan los resultados del análisis. El valor de la constante en cada regresión representa el promedio del grupo de control, mientras que los coeficientes asociados a cada tratamiento indican la diferencia en los resultados entre el tratamiento correspondiente y el grupo de control, es decir, el efecto promedio del tratamiento**.**

Además, los asteriscos señalan los coeficientes que alcanzan significancia estadística en distintos niveles: 0,1 (*), 0,05 (**), 0,01 (***). Estos valores representan la probabilidad de cometer un error tipo I (10%, 5% y 1% respectivamente), es decir, la posibilidad de obtener un falso positivo, donde se rechaza la hipótesis nula aun cuando esta es verdadera. Asimismo, se corrigen posibles problemas de heterocedasticidad mediante el uso de errores estándar robustos (indicados entre paréntesis), lo que permite obtener inferencias más precisas y confiables.

Los Modelos 1 y 3 **(Tabla 4)** muestran que, en términos generales, los tratamientos no influyeron significativamente en la *probabilidad* de clasificar correctamente un correo. Sin embargo, el tratamiento T3 ("*Explicación"*) destacó como una excepción. En la condición experimental T0, en la que los participantes no recibieron información previa al test, este refuerzo incrementó la probabilidad de identificar correctamente un correo de phishing en un 2,3%. La limitada efectividad de los tratamientos podría explicarse, al menos en parte, por el buen desempeño de los grupos de control (T0 y T1), que ya presentaban puntuaciones elevadas en esta variable, reflejadas en la constante de la regresión. Específicamente, ambos grupos de control obtuvieron puntuaciones cercanas a 4 en una escala de 1 a 5, lo que sugiere que los participantes ya poseían un nivel sólido de clasificación antes de la intervención.

Estos resultados podrían sugerir que una campaña estándar, sin refuerzos, basada en información básica para mejorar la detección de correos de *phishing*, puede tener un

10 El grupo T0, dado que no fue expuesto a ninguna campaña, no respondió las preguntas de evaluación de la campaña, sino que pasó directamente a las preguntas de caracterización socioeconómica.

impacto similar al de no proporcionar ninguna información. Esto indicaría que la probabilidad de detectar correos fraudulentos depende principalmente del conocimiento previo de los participantes, más que de la información brindada por la campaña.

Por otra parte, los Modelos 2 y 4 **(Tabla 4)** muestran impactos positivos y significativos de los tratamientos T2 ("*Motivación*"), T3 ("*Explicación*") y T4 ("*Motivación x Explicación")* en el *número de elementos* identificados como señales de *phishing*.

En términos porcentuales, cuando el grupo de control estuvo expuesto a la campaña estándar sin refuerzos (T1), T3 ("*Explicación*") aumentó el indicador asociado al *número de elementos*, en un 8,4% respecto al control, mientras que T2 ("*Motivación)* lo incrementó en un 4,5%. En tanto, cuando el grupo de control estuvo compuesto por participantes que no recibieron información previa al test (T0), los efectos de los tratamientos fueron aún mayores: T3 incrementó el indicador en un 10,3%, y T2 en un 6,4%.
Es importante destacar que los participantes del grupo experimental T3, en promedio, dedicaron 33 segundos más a revisar los correos en comparación con los del grupo de control (T1). Sin embargo, no se encontraron diferencias significativas en el tiempo de revisión respecto a los otros grupos (T2 y T4). Este resultado es alentador, ya que las estafas de *phishing* suelen depender de respuestas impulsivas de las víctimas. El mayor tiempo de revisión observado podría indicar un incremento en la precaución y mayor análisis crítico por parte de los participantes.

Tabla 4: Impactos sobre clasificación de correos

	Experimento 1 (Control T1)		***Experimento 1 (Control T0)***	
	Probabilidad	N° de elementos	Probabilidad	N° de elementos
	(1)	(2)	(3)	(4)
T1: Sin factores			0,032	0,040
			(0,047)	(0,055)
T2: Motivación	-0,010	0,105*	0,022	0,145***
	(0,052)	(0,060)	(0,048)	(0,056)
T3: Explicación	0,060	0,193***	0,092*	0,233***
	(0,051)	(0,060)	(0,047)	(0,056)
T4: Expl. x Mot.	0,032	0,066	0,063	0,106*
	(0,051)	(0,059)	(0,048)	(0,055)
Constante	3,955***	2,308***	3,923***	2,268***
	(0,036)	(0,042)	(0,031)	(0,036)
n	6.555	6.555	8.901	8.901
R2	0,0004	0,002	0,0005	0,002

*Nota: *p<0,1; **p<0,05; ***p<0,01*

Otra forma de visualizar estos resultados es a través de representaciones gráficas, donde se muestra el promedio de cada grupo experimental **(Gráficos 7 al 10).** En estos gráficos, la línea negra representa el valor promedio del grupo de control respectivo, mientras que cada grupo experimental está representado por un punto (su valor promedio) acompañado de un intervalo de confianza del 95%. El efecto del tratamiento se mide como la diferencia vertical entre el punto y la línea negra horizontal.

Se observa que, a este nivel de significancia, solo T3 ("*Explicación*") difiere significativamente del control en el número de elementos detectados cuando el grupo de control es T1 **(Gráfico 8).** En tanto, cuando el grupo de control es T0, tanto T2

("*Motivación*") como T3 ("*Explicación*") presentan diferencias significativas respecto al control **(Gráfico 10).**

Gráfico 7. Impactos en Probabilidad (Control T1)

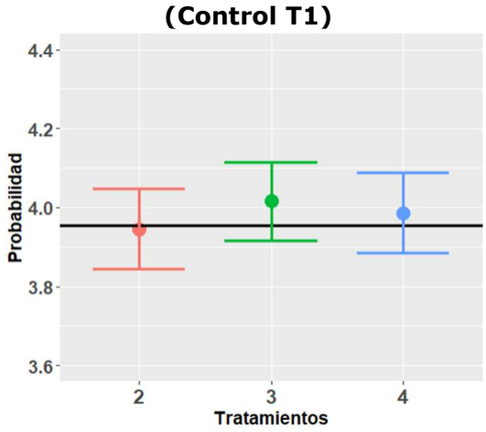


Gráfico 8. Impactos en N° de elementos (Control T1)

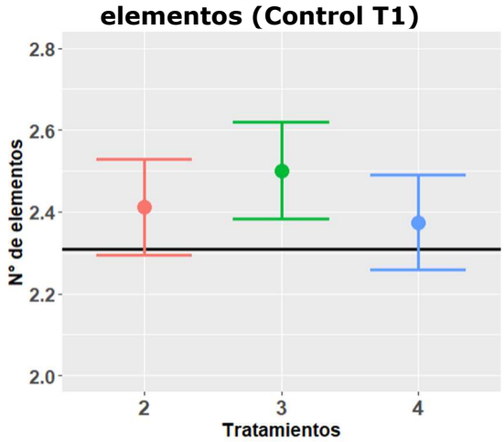


Gráfico 9. Impactos en Probabilidad (Control T0)

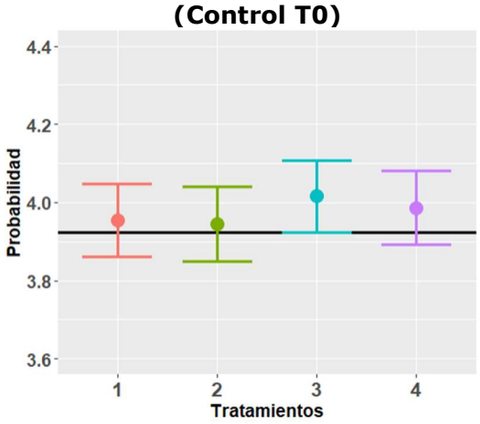
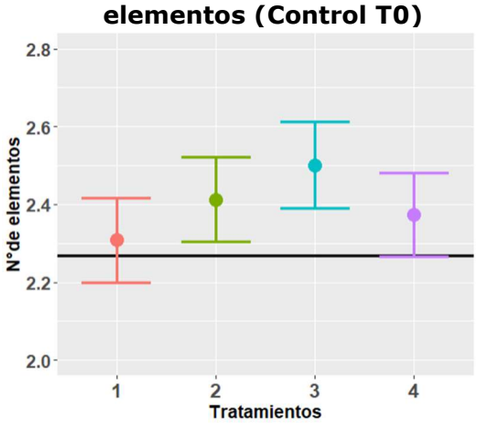


Gráfico 10. Impactos en N° de elementos (Control T0)



Fuente: Elaboración propia.

Experimento 2

El Experimento 2 evalúa cómo distintos elementos comunicacionales, incluidos en las campañas estándar de *phishing*, influyen en los cambios de actitud respecto de las medidas de detección de fraudes en línea ("*actitud*") y en la intención de protegerse ("*intención*"). Estas variables se consideran esenciales para fomentar modificaciones en el comportamiento observable, favoreciendo una mayor disposición a adoptar preventivamente las recomendaciones de seguridad.

Los elementos comunicacionales evaluados incluyen: a) Explicaciones sobre técnicas de ingeniería social ("*Explicación*"), b) los mensajes motivacionales que apelan la autoeficacia y la percepción de gravedad ("*Motivación*") y c) el enfoque de juegos, que permite la inmersión completa de los usuarios en la narrativa del mensaje de la campaña ("*Juego*").

Asimismo, el experimento evaluó tanto las **variables finales** (*actitud* hacia las medidas de seguridad e *intención* de protegerse) como las **variables intermedias**, vinculadas al canal de persuasión, que incidirían en las primeras de acuerdo al Modelo de Procesamiento de la información de McGuire (1969) (ver sección 4.2.5).

Para evaluar el experimento, se utilizaron dos grupos de control:

- **Grupo experimental T5**: compuesto por participantes expuestos a la campaña estándar sin refuerzos, quienes posteriormente respondieron preguntas sobre los resultados finales e intermedios.
- **Grupo experimental T9**: integrado por participantes que recibieron una campaña estándar sin refuerzos, pero con un formato distinto al presentado a los otros grupos experimentales. Esta campaña se asemejaba en diseño y contenido a las utilizadas por instituciones financieras en sus páginas web. Luego, los participantes respondieron preguntas sobre los resultados finales e intermedios.

A la vez, los participantes de todos los grupos experimentales, tras la exposición a los tratamientos, respondieron preguntas relacionadas con los resultados finales (*actitud* e *intención*) y los resultados intermedios asociados al modelo de comunicación y persuasión. Estos últimos incluyen*: Atención* prestada al contenido, *Comprensión* del material presentado, Aceptación de la Campaña (evaluada a través de la *Efectividad percibida*, la *Fuerza argumental percibida*, y el *Gusto general* por la campaña) y Conductores de la Motivación a la protección (medidos mediante la *Gravedad percibida* del fraude, la *Autoeficacia percibida*).

En la **Tabla 5** se presentan los resultados del experimento respecto a las variables "*Actitud*" e "*Intención*". Destaca que todos los tratamientos que incluyen el factor *Juego* (T1, T2, T3 y T4) generan impactos positivos y significativos tanto en actitud como en intención de protección. Por el contrario, los tratamientos sin el factor *Juego* (T5, T6, T7 y T8) no mostraron efectos significativos, con la excepción de T7, que tuvo un impacto en *Actitud* cuando el grupo de control fue T9.

Actitud: En ambos contextos, tanto cuando el grupo de control fue T5 como cuando fue T9, el tratamiento T4 —que combinaba los tres factores (*Juego, Explicación y Motivación*)— incrementó el indicador de actitud en un 6,4% y un 7,6%, respectivamente. Por su parte, el tratamiento T1, que consistió únicamente en incorporar el *Juego* como refuerzo de la campaña estándar, logró incrementos del indicador de un 4,9% y un 6%, respectivamente **(Modelos 1 y 3, Tabla 5).**

Intención: En relación con la variable "intención", el tratamiento T3 (Explicación y Juego) mostró un incremento del índice del 10% y 7,7 % con respecto a los grupos de control T5 y T9, respectivamente. Por su parte, el tratamiento T4, que integra los tres factores de refuerzo, alcanzó un aumento del 9,7% y 7,4%, respectivamente **(Modelos 2 y 4, Tabla 5).**

7

Tabla 5. Impactos en indicadores de Actitud e Intención

	Experimento 2 (Control T5)		***Experimento 2 (Control T9)***	
	Actitud (1)	Intención (2)	Actitud (3)	Intención (4)
T1: Juego	0,191***	0,347***	0,233***	0,262***
	(0,060)	(0,082)	(0,062)	(0,081)
T2: Motivación x Juego	0,129**	0,342***	0,171***	0,257***
	(0,064)	(0,083)	(0,065)	(0,083)
T3: Explicación x Juego	0,157***	0,400***	0,199***	0,315***
	(0,060)	(0,081)	(0,062)	(0,081)
T4: Expl. x Mot. x Juego	0,252***	0,387***	0,294***	0,302***
	(0,059)	(0,081)	(0,061)	(0,081)
T5: Sin factores			0,042	-0,085
			(0,062)	(0,085)
T6: Motivación	-0,035	0,071	0,007	-0,014
	(0,059)	(0,081)	(0,060)	(0,081)
T7: Explicación	0,075	0,028	0,117*	-0,057
	(0,059)	(0,084)	(0,060)	(0,084)
T8: Expl. x Mot.	-0,011	0,107	0,031	0,022
	(0,060)	(0,082)	(0,061)	(0,082)
Constante	3,916***	4,012***	3,874***	4,097***
	(0,043)	(0,060)	(0,045)	(0,060)
n	4.703	4.703	5.262	5.262
R2	0,009	0,014	0,010	0,013

*Nota: *p<0,1; **p<0,05;*** p<0,01*

Los gráficos muestran los resultados de cada grupo experimental en comparación con los grupos de control (Tratamiento 5 y Tratamiento 9, representados por la línea horizontal). En la variable *Actitud*, sobresale el efecto del Tratamiento 4, que integra los tres factores propuestos: *Juego, Explicación y Motivación* **(Gráficos 11 y 12).** Por otro lado, en la variable Intención, se destaca el Tratamiento 3, que combina *Explicación y Juego* **(Gráficos 13 y 14).**

Gráfico 11. Impactos sobre actitud hacia las medidas de seguridad (Control T5)

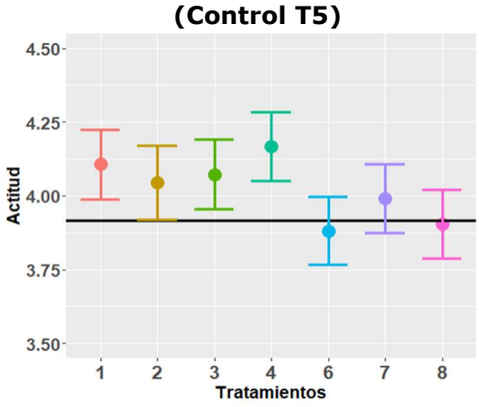


Gráfico 12. Impactos sobre actitud hacia las medidas de seguridad (Control T9)

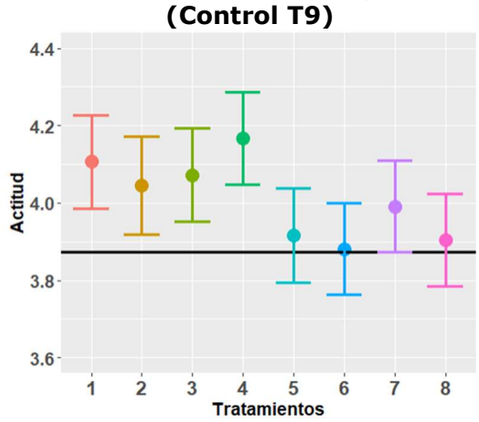


Gráfico 13. Impactos sobre intención de protección (Control T5)

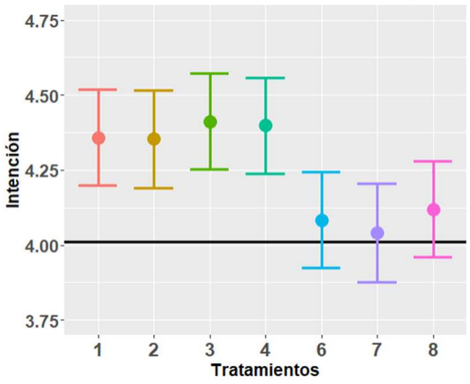
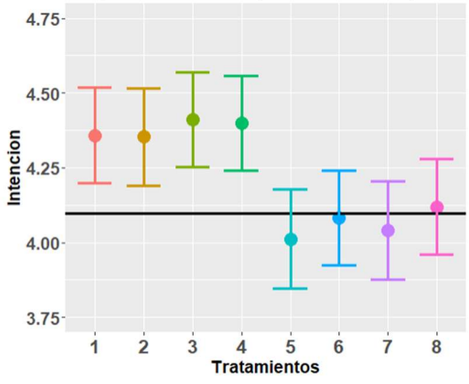


Gráfico 14. Impactos sobre intención de protección (Control T9)



Fuente: Elaboración propia.

Variables Canal de Persuasión

Los resultados sobre las variables intermedias vinculadas al canal de persuasión se presentan en la **Tabla 6.** En términos generales, los tratamientos T1 a T4 (todos ellos incluyen el factor *Juego*) tienden a generar impactos positivos y significativos en la mayoría de las variables, con la excepción de *Gravedad Percibida*, donde los efectos son de menor magnitud en comparación con las demás. Esto podría explicarse por el alto nivel promedio de *Gravedad Percibida* en los grupos de control, lo que sugiere que los participantes ya consideran los fraudes de phishing como un problema grave.

Asimismo, el tratamiento T7 (*Explicación*) muestra un impacto positivo y significativo en *Atención, Fuerza Argumental Percibida, Efectividad Percibida, Gusto General y Autoeficacia*. En contraste, los tratamientos T6 (*Motivación*) y T8 (*Explicación* × *Motivación*) no generan mejoras significativas en ninguna de las variables analizadas.

Estos hallazgos indican que el *Juego* y la *Explicación* son los factores clave para fortalecer las variables asociadas al canal de persuasión.

Es importante destacar que el análisis factorial permitirá identificar con mayor precisión qué factores explican estos impactos, complementando los resultados presentados en esta sección (ver sección 5.3).

Para facilitar la interpretación de los resultados, se emplean representaciones gráficas que muestran el promedio de cada grupo experimental. En estos gráficos, la línea negra horizontal indica el valor promedio del grupo de control correspondiente, mientras que cada grupo experimental se representa mediante un punto, acompañado de un intervalo de confianza del 95%. El impacto del tratamiento se visualiza como la diferencia vertical entre el punto del grupo experimental y la línea negra del grupo de control, lo que permite identificar de manera clara las variaciones en cada variable analizada.

Los gráficos comparan los resultados de los grupos experimentales, resaltando las diferencias en los valores promedio y su nivel de incertidumbre. Esta representación permite observar de manera intuitiva cómo se comportan los distintos tratamientos en relación con los grupos de referencia **(Gráficos 15 a 21).**

Tabla 6. Impactos en variables del canal de persuasión

Experimento 2 (Control T5)

	Atención	Comprensión	Fuerza Argumental Percibida	Efectividad Percibida	Gusto General	Gravedad Percibida	Autoeficacia
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T1: Juego	0,330***	0,075***	0,159***	0,237***	0,190***	0,082*	0,197***
	(0,072)	(0,016)	(0,055)	(0,057)	(0,065)	(0,048)	(0,059)
T2: Motivación x Juego	0,335***	0,052***	0,098*	0,179***	0,167**	0,080	0,170***
	(0,074)	(0,017)	(0,059)	(0,060)	(0,067)	(0,049)	(0,063)
T3: Expl. x Juego	0,308***	0,048***	0,166***	0,255***	0,302***	0,037	0,238***
	(0,073)	(0,017)	(0,056)	(0,057)	(0,064)	(0,050)	(0,060)
T4: Expl. x Motiv. x Jgo.	0,389***	0,056***	0,230***	0,288***	0,277***	0,089*	0,299***
	(0,074)	(0,016)	(0,056)	(0,058)	(0,065)	(0,050)	(0,060)
T6: Motivación	0,006	-0,001	0,038	0,020	-0,021	0,045	-0,014
	(0,073)	(0,017)	(0,054)	(0,055)	(0,063)	(0,048)	(0,060)
T7: Explicación	0,160**	0.017	0,141**	0,146***	0,130**	0,053	0,130**
	(0,073)	(0,017)	(0,055)	(0,055)	(0,062)	(0,048)	(0,061)
T8: Expl. x Motivación	0,038	0.021	0,085	0,071	-0,005	-0,019	0,035
	(0,074)	(0,017)	(0,055)	(0,056)	(0,063)	(0,050)	(0,060)
Constante	3,564***	0,802***	3,872***	3,390***	3,428***	4,589***	3,819***
	(0,053)	(0,012)	(0,039)	(0,040)	(0,045)	(0,036)	(0,043)
N	4.703	4.703	4.703	4.703	4.703	4.703	4.703
R2	0,015	0,009	0,005	0,011	0,012	0,002	0,011

*Nota: *p<0,1; **p<0,05; ***p<0,01*

Gráfico 15: Atención

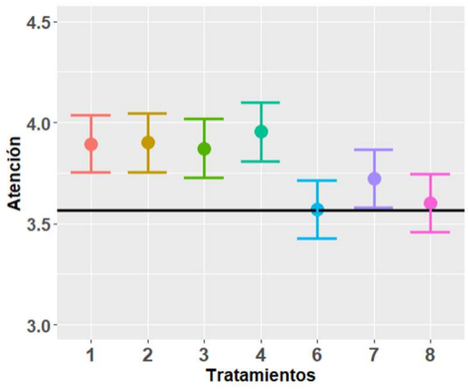


Gráfico 16: Comprensión

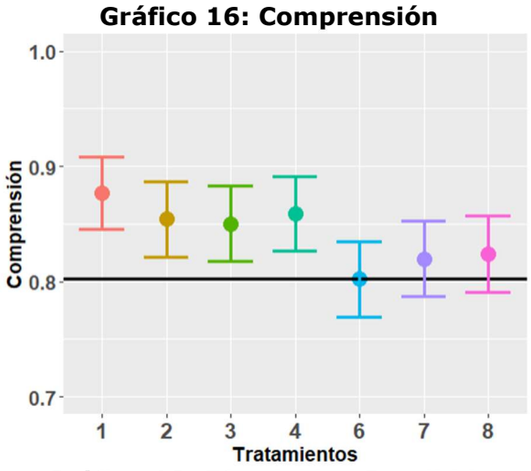


Gráfico 17: Fuerza Argumental Percibida

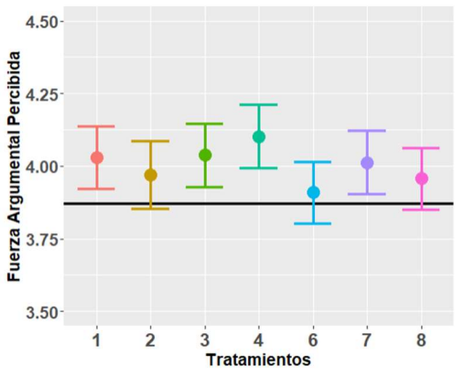


Gráfico 18: Efectividad Percibida

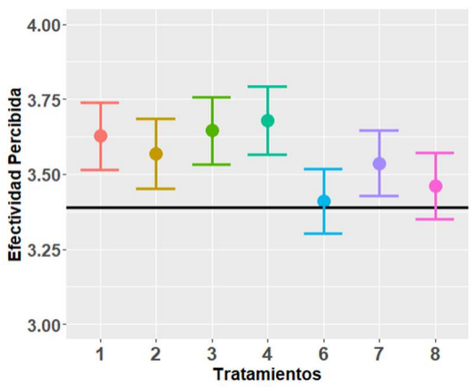


Gráfico 19: Gusto general

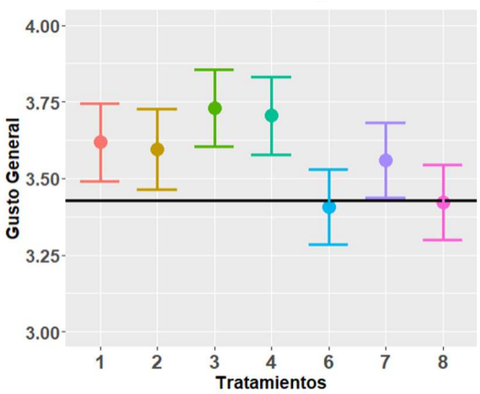


Gráfico 20: Autoeficacia Percibida

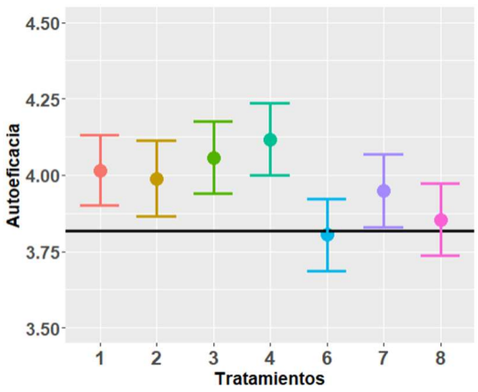
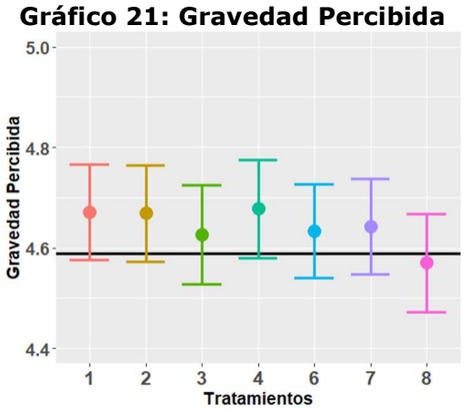


Gráfico 21: Gravedad Percibida



5.2.3 Análisis de Robustez

En esta sección se evalúa la solidez de los efectos de los tratamientos estimados previamente. El análisis de robustez tiene como objetivo demostrar que los resultados del estudio son consistentes y no dependen críticamente ni de la muestra ni de las especificaciones funcionales utilizadas para estimar dichos efectos, sino que se mantienen consistentes al aplicar métodos alternativos.

Para ello se emplean tres enfoques:

- **Inclusión de covariables**: Se añaden variables de control para ajustar posibles diferencias entre grupos. Esto permite reducir sesgos incluso cuando el análisis de balance previo no detectó diferencias significativas.

- **Ampliación de la muestra:** Se amplió la muestra, casi duplicándola, al incluir datos de encuestas incompletas. Esto permite evaluar la estabilidad de los resultados ante posibles sesgos por desgaste (*attrition bias*), el cual se define como un error sistemático que surge cuando se pierden participantes durante el transcurso de un estudio. Si las características o resultados de los que abandonan el estudio difieren de los que lo completan, los hallazgos pueden estar sesgados y no ser representativos de la población inicial, afectando así la validez interna y externa de la investigación (Higgins et al., 2019).

- **Pruebas no paramétricas:** Cuando las variables dependientes son ordinales, se recomienda utilizar pruebas no paramétricas, ya que no requieren asumir normalidad ni intervalos equidistantes entre categorías. Estas pruebas permiten comparar diferencias entre grupos y realizar inferencias de manera más adecuada. En este contexto, se aplicó la prueba de Kruskal-Wallis, que evalúa si las distribuciones de los grupos difieren con base en los rangos de los datos. Para identificar qué grupos presentan diferencias significativas, se emplearon pruebas post hoc de Dunn, las cuales determinan si las distribuciones de los grupos difieren en ubicación y forma, ajustando los valores p mediante la corrección de Bonferroni para minimizar el riesgo de cometer errores tipo I (Kraska-Miller, 2013).

En términos generales, los resultados se consideran robustos si los coeficientes estimados mantienen su magnitud, signo y significancia. Además, los métodos empleados pueden mejorar la precisión de las estimaciones: la incorporación de covariables relevantes reduce el error de la regresión, mientras que un mayor tamaño muestral aumenta el poder estadístico, especialmente cuando el desgaste es aleatorio. Asimismo, el uso de pruebas no paramétricas es útil cuando no se cumplen los supuestos de normalidad o de medición en escala de intervalo, permitiendo obtener inferencias más confiables y reduciendo posibles sesgos derivados de la distribución de los datos o de la elección de la prueba estadística aplicada (Angrist & Pischke, 2009; Wooldridge, 2010; Cameron & Trivedi, 2005).

Experimento 1: Inclusión de covariables

Se emplearon las siguientes covariables, explicadas en la sección 4.2.6, algunas del tipo *dummy* y otras continuas. En el caso de las variables *dummy*, para evitar la multicolinealidad perfecta, es necesario excluir al menos una de las categorías *dummy*, ya que su efecto queda capturado por el intercepto de la regresión.

Las variables *dummy* incluidas son: género femenino; tres de los cuatro tramos de edad (30–44, 45–59 y 60+); tres de los cinco niveles de educación (Técnica, Universitaria y Postgrado); y cinco de los seis tramos de ingreso (250M–500M; 501M–900M; 900M–1,5MM; 1,5MM–2,0MM; 2,0MM+). Además, se incluyeron las variables *dummy* "*móvil*" (encuesta respondida desde un teléfono móvil) y "*víctima*" (ha sido víctima de fraude durante el último año).

En el caso de las variables continuas, se incluyeron dos: "*experiencia"* (en el uso de internet) y "*necesidad de cognición"* (la propensión a disfrutar de la actividad de pensar). Los modelos 1 y 3 de la **tabla 7**, presenta los resultados originales (sin controles), mientas que los modelos 2 y 4 muestra las estimaciones que incorporan las variables de control. Se evidencia que los efectos de T2 (*Motivación*) y T3 (*Explicación*) sobre la *probabilidad* de clasificar correctamente un correo y sobre el *número de elementos* detectados son robustos; es decir, se mantienen en magnitud, signo y significancia.

Tabla 7: Impactos sobre clasificación de correos (inclusión de covariables)
Experimento 1 (Control T1)

	Probabilidad		N° de elementos	
	(1)	(2)	(3)	(4)
T2: Motivación	-0,010	-0,031	0,105*	0,104*
	(0,052)	(0,051)	(0,060)	(0,060)
T3: Explicación	0,060	0,041	0,193***	0,171***
	(0,051)	(0,050)	(0,060)	(0,060)
T4: Exp. x Mot.	0,032	0,043	0,066	0,065
	(0,051)	(0,050)	(0,059)	(0,059)
Constante	3,955***	2,978***	2,308***	1,515***
	(0,036)	(0,157)	(0,042)	(0,191)
Controles	No	Sí	No	Sí
N	6.555	6.555	6.555	6.555
R2	0,0004	0,037	0,002	0,015
Adj. R2	-0,0001	0,035	0,001	0,012

*Nota: *p<0,1; **p<0,05; ***p<0,01*

Experimento 1: Muestra ampliada

En el caso de los impactos de los tratamientos sobre la *probabilidad* de clasificar correctamente los correos y el *número de elementos* de *phishing* detectados, se realizó el análisis mediante dos enfoques:

- **Inclusión de encuestas incompletas y clasificación incompleta**: Se consideraron todas aquellas encuestas en las que los participantes clasificaron al menos el primer correo, teniendo en cuenta que cada encuestado debía clasificar tres correos y que el orden de presentación fue idéntico para todos. Los resultados obtenidos a partir de esta muestra se presentan en los **Modelos 2 y 5 de la Tabla 8.**

Dado que el primer correo (correspondiente a un *phishing* de Banco Estado) está sobrerrepresentado en esta muestra y que los participantes tardaron más en responderlo en comparación con los otros correos (mediana: 107 segundos para el primero, 72 para el segundo y 56 para el tercero), esta estimación no es completamente comparable con la muestra original. Para corregir esta diferencia, se presenta el segundo enfoque.

- **Inclusión de encuestas incompletas y clasificación completa**: Se incluyeron únicamente las encuestas que, aunque incompletas en otros aspectos, contenían las clasificaciones de los tres correos. Los resultados de esta muestra se presentan en los **Modelos 3 y 6 de la Tabla 8.**

Tabla 8: Impactos sobre clasificación de correos (muestra ampliada)

	Experimento 1 (Control T5)					
	Probabilidad			N° de elementos		
	(1)	(2)	(3)	(4)	(5)	(6)
T2: Motivación	-0,010	0,024	0,011	0,105*	0,070	0,085*
	(0,052)	(0,040)	(0,042)	(0,060)	(0,046)	(0,047)
T3: Explicación	0,060	0,024	0,027	0,193***	0,148***	0,150***
	(0,051)	(0,041)	(0,042)	(0,060)	(0,046)	(0,048)
T4: Expl. x Mot.	0,032	0,046	0,024	0,066	0,009	0,017
	(0,051)	(0,041)	(0,042)	(0,059)	(0,046)	(0,047)
Constante	3,955***	3,842***	3,876***	2,308***	2,401***	2,343***
	(0,036)	(0,028)	(0,029)	(0,042)	(0,032)	(0,033)
N	6.555	11.673	10.596	6.555	11.673	10.596
R2	0,0004	0,0001	0,00005	0,002	0,001	0,001

*Nota: *p<0,1; **p<0,05; ***p<0,01*

Se observa que, en ambos casos, la muestra se amplió significativamente, con una diferencia reducida entre las muestras ampliadas (11.673 vs. 10.596 encuestas). Esto sugiere que el abandono durante el juego fue bajo, aproximadamente un 10%, lo que indica una alta disposición de los participantes a involucrarse en un entorno de juego en condiciones reales. Desde la perspectiva del Modelo de Probabilidad de Elaboración (EELM, por sus siglas en inglés), la presentación de una narrativa persuasiva a través de un juego incrementa la implicación del individuo, sumergiéndolo en la historia y reduciendo su nivel de crítica hacia el contenido. Como resultado, disminuyen los contraargumentos y la resistencia al mensaje, favoreciendo así la efectividad de la persuasión (Shrum, 2004; Slater & Rouner, 2002).

En cuanto a los coeficientes estimados, se evidencia que el efecto promedio de T3 (*Explicación*) sobre el *número de elementos* detectados es robusto, mientras que para T2 (*Motivación*) dejó de ser significativa en el Modelo 5, lo que podría explicarse por la menor comparabilidad de esa estimación **(Tabla 8).**

Experimento 1: Pruebas no Paramétricas

A continuación, se presentan los resultados del test de Kruskal-Wallis, una prueba no paramétrica equivalente al ANOVA de una vía, utilizada para comparar más de dos grupos. Asimismo, se reportan los resultados de la prueba de Dunn, un análisis post hoc que permite identificar qué grupos presentan diferencias significativas. Para controlar el riesgo de errores tipo I, los valores de p han sido ajustados mediante la corrección de Bonferroni.

Los resultados del Test de Kruskal-Wallis presentados en la **tabla 9** indican que no hay diferencias significativas en la *Probabilidad* de clasificar correctamente los correos entre los grupos analizados (T1 a T4) (H = 2,63, p = 0,45), ya que el p-valor es mucho mayor a 0,05, lo que indica que los tratamientos no generan un impacto significativo en esta variable.

Por otro lado, en la variable *Número de Elementos*, sí se encuentra una diferencia significativa entre los grupos (H = 9,84, p = 0,02), lo que sugiere que al menos un grupo

presenta un impacto distinto en la cantidad de elementos identificados. Dado que el p-valor es menor a 0,05, se concluye que al menos uno de los tratamientos aplicados afecta significativamente el número de elementos detectados en la clasificación de correos.

Tabla 9: Impactos sobre clasificación de correos (Test Kruskal-Wallis)

	H	p-value
Probabilidad	2,63	0,45
N°Elementos	9,84	0,02

A continuación, se presentan los resultados de las pruebas post hoc de Dunn con ajuste de Bonferroni, utilizada para identificar los grupos que presentan diferencias significativas. Los resultados reportados en la Tabla 10, muestran que solo la comparación entre los grupos T1 y T3 muestra un efecto estadísticamente significativo en el *número de elementos detectados*, mientras que las demás comparaciones no presentan diferencias significativas en ninguna de las variables analizadas.

Tabla 10: Impactos sobre clasificación de correos (Prueba Post Hoc Dunn)

	1-2 Z score	1-3 Z score	1-4 Z score
Probabilidad	-0,09	-1,34	-1,00
N° de elementos	-1,78	-3,10**	-1,31

<u>Resultados generales de las pruebas de Robustez respecto a la clasificación correcta de correos entre Fraudulentos y legítimos</u>

Los resultados de las pruebas de robustez confirman que incluir *explicaciones* sobre técnicas de ingeniería social o *mensajes motivacionales*, no produce un incremento estadísticamente significativo en la *probabilidad* de clasificar correctamente los correos electrónicos (entre fraudulentos y legítimos).

Por otro lado, todas las pruebas de robustez, tanto las paramétricas como las no paramétricas, evidencian un efecto positivo y estadísticamente significativo del tratamiento T3 en el *número de elementos* identificados. Es decir, el tratamiento que incorpora *explicaciones* sobre ingeniería social, como un refuerzo a la campaña estándar, aumenta la cantidad de elementos detectados por los participantes.

En cuanto al tratamiento T2 (que incluye *mensajes motivacionales*), los resultados muestran un efecto positivo y estadísticamente significativo sobre el *número de elementos* identificados según las pruebas de robustez paramétricas; sin embargo, este efecto no se mantiene en la prueba no paramétrica.

Experimento 2: Inclusión de covariables

En la **tabla 11** se presenta el análisis de robustez para el Experimento 2, utilizando las mismas covariables empleadas en la prueba de robustez del Experimento 1. Se evalúan dos variables dependientes: la *actitud* hacia las medidas de seguridad y la *intención* de protegerse. En todos los casos, las magnitudes, los signos y las significancias se mantienen consistentes, lo que confirma la robustez de los resultados anteriores.

Tabla 11: Impactos en indicadores de Actitud e Intención (inclusión de covariables)

Experimento 2 (Control T5)

	Actitud		Intención	
	(1)	(2)	(3)	(4)
T1: Juego	0,191***	0,189***	0,347***	0,349***
	(0,060)	(0,059)	(0,082)	(0,081)
T2: Motivación x Juego	0,129**	0,137**	0,342***	0,359***
	(0,064)	(0,062)	(0,083)	(0,082)
T3: Explicación x Juego	0,157***	0,154***	0,400***	0,400***
	(0,060)	(0,059)	(0,081)	(0,080)
T4: Expl. x Mot. x Juego	0,252***	0,259***	0,387***	0,396***
	(0,059)	(0,058)	(0,081)	(0,080)
T6: Motivación	-0,035	-0,026	0,071	0,079
	(0,059)	(0,058)	(0,081)	(0,080)
T7: Explicación	0,075	0,072	0,028	0,034
	(0,059)	(0,058)	(0,084)	(0,082)
T8: Expl. x Mot.	-0,011	-0,008	0,107	0,109
	(0,060)	(0,058)	(0,082)	(0,081)
Constante	3,916***	2,557***	4,012***	2,634***
	(0,043)	(0,134)	(0,060)	(0,180)
Controles	No	Sí	No	Sí
N	4.703	4.703	4.703	4.703
R2	0,009	0,056	0,014	0,048
Adj. R2	0,008	0,051	0,012	0,043

*Nota: *p<0,1; **p<0,05; ***p<0,01*

Experimento 2: Muestra ampliada

Para evaluar la robustez de los efectos sobre las variables de *actitud* e *intención*, se utilizaron todas las respuestas que contaban con datos para el indicador correspondiente. En ambos casos—tanto para la *actitud* como para la *intención*—la muestra ampliada comprende 8.067 respuestas. Con respecto a los impactos, se observa que las magnitudes, los signos y las significancias se mantienen, lo que indica que los resultados son robustos. Además, se detecta un nuevo efecto significativo en el grupo experimental T7 (*Explicación*) sobre la *actitud*, probablemente debido al mayor tamaño muestral, que incrementa el poder estadístico al reducir el error estándar, y a un efecto estimado más elevado para este tratamiento **(Tabla 12).**

Tabla 12: Impactos en indicadores de Actitud e Intención (muestra ampliada)

	Experimento 2 (Control T5)			
	Actitud		Intención	
	(1)	(2)	(3)	(4)
T1: Juego	0,191***	0,235***	0,347***	0,354***
	(0,060)	(0,044)	(0,082)	(0,057)
T2: Motivación x Juego	0,129**	0,178***	0,342***	0,329***
	(0,064)	(0,047)	(0,083)	(0,059)
T3: Expl. x Juego	0,157***	0,201***	0,400***	0,385***
	(0,060)	(0,044)	(0,081)	(0,057)
T4: Expl. x Mot. x Juego	0,252***	0,325***	0,387***	0,392***
	(0,059)	(0,043)	(0,081)	(0,058)
T6: Motivación	-0,035	-0,012	0,071	-0,024
	(0,059)	(0,041)	(0,081)	(0,057)
T7: Explicación	0,075	0,112***	0,028	0,010
	(0,059)	(0,041)	(0,084)	(0,057)
T8: Expl. x Motivación	-0,011	0,010	0,107	0,055
	(0,060)	(0,042)	(0,082)	(0,056)
Constante	3,916***	3,802***	4,012***	3,982***
	(0,043)	(0,025)	(0,060)	(0,034)
n	4.703	8.067	4.703	8.067
R2	0,009	0,012	0,014	0,016

*Nota: *p<0,1; **p<0,05; ***p<0,01*
.

Experimento 2: Pruebas no Paramétricas

La **Tabla 13** presenta los resultados del Test de Kruskal-Wallis, evidenciando diferencias estadísticamente significativas en la distribución de la mayoría de las variables analizadas, excepto en *gravedad percibida* (p = 0.06), donde no se encontró suficiente evidencia para afirmar diferencias entre los grupos. La mayor diferencia entre grupos se observó en *intención* (H = 123, p = 0.000), seguida de *atención* (H = 77, p = 0.000) y *gusto general* (H = 71, p = 0.000). Otras variables, como *actitud, comprensión, fuerza argumental percibida, efectividad percibida y autoeficacia percibida*, también mostraron diferencias significativas (p < 0.05). Estos resultados sugieren que las percepciones y respuestas de los grupos varían significativamente en la mayoría de las variables analizadas.

Tabla 13: Impactos en indicadores de Actitud e Intención y variables del canal de persuasión (Test Kruskal- Wallis)

	H	p-value
Actitud	52,18	0
Intención	118,13	0
Atención	73,09	0
Comprensión	42,46	0
Fuerza Argumental Percibido	28,62	0,00017
Efectividad Percibida	60,16	0
Gusto General	60,54	0
Gravedad Percibida	12,75	0,0784
Autoeficacia Percibida	56,94	0

A continuación, se presentan los resultados de las pruebas post hoc de Dunn con ajuste de Bonferroni, utilizada para identificar los grupos que presentan diferencias significativas.

Los resultados reportados en la **Tabla 14**, revelan diferencias significativas en varias variables analizadas, con las mayores discrepancias observadas en *intención, atención y efectividad percibida*, que presentan múltiples comparaciones altamente significativas ($p < 0.001$). En particular, la variable *intención* presenta las diferencias más marcadas, con valores Z elevados (5,46, 5,59, 6,37 y 5,88), indicando una separación clara entre grupos. Asimismo, *gusto general* y *autoeficacia percibida* muestran diferencias significativas en algunos grupos, mientras que *actitud* y *fuerza argumental percibida* presentan diferencias más puntuales. En contraste, *gravedad percibida* no muestra diferencias significativas, lo que coincide con los resultados del Test de Kruskal-Wallis ($p = 0,06$), indicando que los grupos tienen percepciones similares sobre esta variable. En general, los hallazgos sugieren que las percepciones y respuestas de los participantes varían significativamente en la mayoría de las variables, especialmente entre los niveles 1-5 y 4-5, donde se identificaron las diferencias más marcadas.

Tabla 14: Impactos en indicadores de Actitud e Intención y variables del canal de persuasión (Pruebas Post Hoc Dunn)[11]

	1-5 Z scr	2-5 Z scr	3-5 Z scr	4-5 Z scr	6-5 Z scr	7-5 Z scr	8-5 Z scr
Actitud	3,55**	2,97*	2,55	4,34***	0,75	-1,19	0,22
Intención	5,50***	5,62***	6,41***	5,91***	-0,57	-0,35	-1,32
Atención	4,15***	4,64***	4,06***	5,46***	0,03	-2,21	-0,39
Comprensión	4,51***	3,01*	2,82	3,21**	0,41	-1,01	-1,41
Fuerza Argumental Percibido	2,68	2,34	2,96*	4,35***	-0,59	-3,00*	-1,58
Efectividad Percibida	4,35***	3,61***	4,61***	5,33***	-0,31	-2,73	-1,28
Gusto General	3,14**	2,96*	4,82***	4,53***	0,22	-2,19	0,07
Gravedad Percibida	1,83	1,72	0,83	2,03	-0,82	-0,79	0,74
Autoeficacia Percibida	2,97*	3,33**	4,05***	5,40***	-0,01	-2,75	-0,59

*Nota: *p<0,1; **p<0,05; *** p<0,01*

Resultados generales de las pruebas de Robustez respecto a los Cambios en la Actitud hacia las medidas de seguridad e Intención de Protección:

Las pruebas de robustez confirman que todos los tratamientos que incorporan el factor Juego (T1, T2, T3 y T4) generan efectos positivos en la *intención de protección* y tienen significancia estadística ($p < 0,01$). En contraste, los tratamientos sin el factor Juego (T5, T6, T7 y T8) no presentan efectos estadísticamente significativos en esta variable.

En cuanto a la variable *actitud*, las pruebas paramétricas evidencian la significancia estadística de los tratamientos con el factor *Juego* (T1 a T4). Sin embargo, en la prueba no paramétrica, el tratamiento T3 no alcanza significancia estadística. Por otra parte, al ampliar la muestra, el tratamiento T7 (*Explicación*) evidencia un efecto positivo y estadísticamente significativo en la *actitud*.

En conclusión, los hallazgos refuerzan la importancia del *Juego* como componente clave para fomentar la *intención de protección* y mejorar la *actitud* hacia la seguridad. Asimismo, resaltan la importancia de combinar el *Juego*, la *Explicación* y la *Motivación* para optimizar el impacto de las campañas contra el phishing en el comportamiento preventivo de los participantes.

[11] En el anexo se presentan los resultados completos de las Pruebas Post Hoc Dunn, para las variables actitud e intención.

5.2.4 Análisis de Heterogeneidad

En esta sección se presenta un análisis de heterogeneidad de los resultados, cuyo objetivo es determinar si los efectos de los tratamientos difieren entre subgrupos de la población. En particular, se examina la variación de los impactos según género (*hombres* versus *mujeres*) y grupo etario (*adultos-jóvenes,* de 18 a 44 años; y *adultos-mayores*, de 45 años o más). Las **tablas 15 y 16** muestran el tamaño muestral correspondiente a cada grupo analizado: la primera tabla corresponde al Experimento 1, y la segunda al Experimento 2.

Cabe destacar que, a pesar de que los tamaños muestrales del Experimento 2 son menores, éstos permiten detectar impactos de entre 0,21 y 0,26 (Cohen's d), es decir, efectos relativamente pequeños, lo que garantiza un alto poder estadístico en el análisis de heterogeneidad.

Tabla 15. Exp 1: Tamaño muestral de subgrupos en clasificación de correos

	Grupo tratamiento			
Subgrupo	**1**	**2**	**3**	**4**
Hombres	912	885	819	840
Mujeres	762	711	798	774
Adultos Mayores (45+)	810	759	780	807
Adultos Jóvenes (18-44)	873	855	852	819

Fuente: Elaboración propia.

Tabla 16. Exp 2: Tamaño muestral de subgrupos en evaluación de campañas

	Grupo tratamiento							
Subgrupos	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Hombres	304	295	273	280	300	311	323	308
Mujeres	254	237	266	258	302	348	299	309
Adultos Mayores (45+)	270	253	260	269	299	340	337	315
Adultos Jóvenes (18-44)	291	285	284	273	306	324	290	307

Fuente: Elaboración propia.

Los impactos se estimaron mediante una ecuación de regresión en la que las *dummies* de los tratamientos interactúan con la *dummy* del subgrupo (género o edad). Por ejemplo, si usamos *H* como variable de heterogeneidad (*dummy* de género o grupo etario), y consideramos sólo un grupo de tratamiento, *T*, para simplificar, la ecuación de regresión sería la siguiente:

$$y = \beta_0 + \beta_1 T + \beta_2 H + \beta_3 TH + \varepsilon$$

El efecto del tratamiento *T* sería $\beta_1 + \beta_3 H$, que depende de *H*. Si el coeficiente β_3 es significativo, entonces se concluye que el efecto de *T* varía con el subgrupo.

Experimento 1: Heterogeneidad según género

La **Tabla 17** muestra los resultados del Experimento 1, en el que se evaluó tanto la *probabilidad* de clasificar correctamente un correo (entre *phishing* y legítimo) como la identificación del *número de elementos* manipuladores presentes en los correos, analizados desde la perspectiva de género (donde H = 1 para el género femenino y 0 para el masculino).

En cuanto al *número de elementos* **(Modelo 2, Tabla 17),** se observa que los efectos principales no son estadísticamente significativos; sin embargo, los tratamientos T2 (*motivación*) y T3 (*explicación*) muestran efectos positivos y significativos al interactuar con la variable dummy de *mujeres*. Esto evidencia la existencia de efectos heterogéneos por género, siendo mayores y exclusivos para las mujeres. Por otro lado, en el Modelo 1, para la variable *probabilidad* no se encontraron resultados significativos.

Tabla 17. Efectos heterogéneos en clasificación de correos (género)

	Experimento 1 (Control T1)	
	Probabilidad (1)	N° de elementos (2)
T2: Motivación	0,021	0,002
	(0,071)	(0,080)
T3: Explicación	0,033	0,096
	(0,072)	(0,082)
T4: Expl. x Mot.	0,058	-0,008
	(0,071)	(0,080)
Mujeres	-0,062	-0,118
	(0,072)	(0,084)
T2': Mot. x Mujeres	-0,077	0,201*
	(0,105)	(0,121)
T3': Expl. x Mujeres	0,051	0,207*
	(0,103)	(0,121)
T4': Expl. x Mot.x Mujeres	-0,044	0,138
	(0,103)	(0,118)
Constante	3,980***	2,362***
	(0,050)	(0,057)
N	6.501	6.501
R^2	0,001	0,002

Nota: *p<0,1; **p<0,05; *** p<0,01

Experimento 1: Heterogeneidad según edad

Al evaluar los efectos heterogéneos según grupo etario, se hizo interactuar los tratamientos con una variable *dummy* para los *adultos-jóvenes*, asignándose H = 1 a los participantes menores de 45 años y H = 0 a los *adultos-mayores* (45 años o más). Los resultados del Experimento 1, respecto a la clasificación de correos, son presentados en la **Tabla 18,** y muestra que no se detectaron efectos heterogéneos, ya que los coeficientes de las interacciones no son significativos. Sin embargo, destaca que la dummy para *adultos-jóvenes* (menores de 45 años) es positiva y significativa para la *probabilidad* de clasificar un correo correctamente, lo que indica que, en promedio, los *adultos-jóvenes* tienen mayor probabilidad de clasificar correctamente los correos en comparación con los participantes de mayor edad.

Tabla 18. Efectos heterogéneos en clasificación de correos (edad)

	Experimento 1 (Control T1)	
	Probabilidad (1)	N° de elementos (2)
T2: Motivación	0,053	0,082
	(0,079)	(0,084)
T3: Explicación	0,064	0,192**
	(0,077)	(0,085)
T4: Expl. x Mot.	0,023	0,074
	(0,078)	(0,082)
Menores 45 años	0,237***	0,086
	(0,072)	(0,084)
T2': Mot. x Menores 45 años	-0,124	0,041
	(0,104)	(0,119)
T3': Expl. x Menores 45 años	-0,008	0,002
	(0,102)	(0,121)
T4': Expl. x Mot. x Menores 45 años	0,024	-0,013
	(0,103)	(0,118)
Constante	3,832***	2,263***
	(0,055)	(0,059)
N	6.555	6.555
R2	0,006	0,002

Nota: *p<0,1; **p<0,05; *** p<0,01

Experimento 2: Heterogeneidad según género

Respecto a los impactos de los tratamientos sobre las variables de *actitud* e *intención*, no se evidencian efectos heterogéneos por género, ya que ninguna de las interacciones con la *dummy* de género femenino resultó significativa. Esto indica que los efectos sobre la *actitud* hacia las medidas de seguridad y la *intención* de protegerse son independientes del género de los participantes **(Tabla 19).**

Tabla 19. Efectos heterogéneos en indicadores de campaña (género)

	Experimento 2 (Control T5)	
	Actitud (1)	Intención (2)
T1: Juego	0,178**	0,412***
	(0,080)	(0,117)
T2: Mot. x Juego	0,063	0,372***
	(0,088)	(0,118)
T3: Expl. x Juego	0,100	0,322***
	(0,086)	(0,124)
T4: Expl. x Mot. x Juego	0,197**	0,409***
	(0,084)	(0,119)
T6: Motivación	-0,018	0,029
	(0,084)	(0,121)
T7: Explicación	0,054	0,021
	(0,083)	(0,121)
T8: Exlp. x Mot.	0,004	0,177
	(0,083)	(0,118)
Mujeres	-0,059	0,176
	(0,086)	(0,121)
T1': Juego x Mujeres	0,029	-0,111
	(0,121)	(0,164)
T2': Mot. x Juego x Mujeres	0,147	-0,041
	(0,127)	(0,168)
T3': Expl. x Juego x Mujeres	0,109	0,169
	(0,120)	(0,163)
T4': Expl. x Mot. x Juego x Mujeres	0,114	-0,036
	(0,119)	(0,163)
T6': Mot. x Mujeres	-0,024	0,077
	(0,118)	(0,164)
T7': Expl. x Mujeres	0,063	0,050
	(0,117)	(0,168)
T8': Expl. x Mot. x Mujeres	-0,024	-0,147
	(0,120)	(0,165)
Constante	3,945***	3,920***
	(0,060)	(0,087)
N	4.667	4.667
R2	0,010	0,019

Nota: *p<0,1; **p<0,05; *** p<0,01

Experimento 2: Heterogeneidad según edad

El análisis de heterogeneidad por edad revela que los *adultos- jóvenes* (<45 años) tienden a tener una *actitud* menos favorable hacia las medidas de seguridad y una menor *intención* de protegerse. No obstante, algunos tratamientos generan efectos diferenciados en este grupo. En particular, la combinación de *explicación y juego* (T3') impacta positivamente en su *actitud*, mientras que la interacción entre explicación, *motivación* y *juego* (T4') influye significativamente en su *intención* de protección. Es

relevante destacar que, aunque T4 (*Explicación, Motivación y Juego*) no tiene un impacto significativo en la *intención* de protección en la muestra general, su interacción con los *adultos jóvenes* (T4') sí lo tiene, lo que indica que su efecto se concentra en este grupo. Además, se observan efectos significativos en la *intención* de protección con la interacción entre *juego* y *adultos jóvenes* (T1') y la combinación de *explicación*, *motivación* y *adultos jóvenes* (T8'). Estos hallazgos sugieren que, aunque los *adultos jóvenes* presentan menor disposición inicial hacia la seguridad, estrategias que combinan *juego*, *motivación* y *explicación* pueden mejorar tanto su *actitud* como su *intención* de protegerse **(Tabla 20).**

Tabla 20. Efectos heterogéneos en indicadores de campaña (edad)

	Experimento 2 (Control T5)	
	Actitud (1)	Intención (2)
T1: Juego	0,161*	0,200*
	(0,084)	(0,117)
T2: Mot. x Juego	0,132	0,215*
	(0,086)	(0,119)
T3: Expl. x Juego	0,034	0,367***
	(0,086)	(0,111)
T4: Expl. x Mot. x Juego	0,173**	0,101
	(0,080)	(0,121)
T6: Motivación	-0,116	-0,040
	(0,082)	(0,114)
T7: Explicación	0,045	-0,048
	(0,079)	(0,115)
T8: Expl. x Mot.	-0,034	-0,048
	(0,080)	(0,117)
Menores 45 años	-0,211**	-0,248**
	(0,085)	(0,120)
T1': Juego x Menores 45 años	0,063	0,289*
	(0,120)	(0,163)
T2': Mot. x Juego x Menores 45 años	0,004	0,251
	(0,126)	(0,167)
T3': Expl. x Juego x Menores 45 años	0,242**	0,072
	(0,120)	(0,162)
T4': Expl. x Mot. x Juego x Menores 45 años	0,157	0,567***
	(0,118)	(0,162)
T6': Mot. x Menores 45 años	0,159	0,219
	(0,117)	(0,162)
T7': Expl. x Menores 45 años	0,046	0,142
	(0,117)	(0,168)
T8': Expl. x Mot. x Menores 45 años	0,042	0,309*
	(0,119)	(0,164)
Constante	4,023***	4,137***
	(0,057)	(0,083)
N	4.703	4.703
R2	0,014	0,017

Nota: *p<0,1; **p<0,05; *** p<0,01

Fuente: elaboración propia.

5.3 Resultados del Modelo basado en Factores

A continuación, se presentan los resultados del análisis del ANOVA factorial, que permite examinar efectos principales (impacto individual de cada factor, cuando los otros factores se encuentran ausentes) y efectos de interacción (cómo un factor influye en la variable dependiente dependiendo de otro factor).

Experimento 1

El experimento evaluó el impacto de los factores "*Explicación*" y "*Motivación*" en la *probabilidad* de clasificar correctamente correos y en la *cantidad de elementos detectados.*

El análisis ANOVA reveló que, aunque estos factores no tuvieron efectos significativos en la *probabilidad* de detección correcta, lo que podría atribuirse a un desempeño uniformemente alto en todos los grupos, incluido el control, sí se identificaron efectos significativos en el número de elementos detectados. En particular, el factor "*Explicación*" tuvo un efecto positivo significativo ($p < 0,1$), mientras que la interacción entre "*Explicación y Motivación*" mostró un efecto altamente significativo, pero negativo ($p < 0,01$). Estos hallazgos sugieren que, si bien la *explicación* por sí sola puede mejorar la identificación de elementos en correos fraudulentos, su combinación con *motivación* puede generar un efecto adverso, posiblemente por una sobrecarga cognitiva o interferencias en el procesamiento de la información (**Tabla 21**).

Tabla 21. ANOVA: significancia de factores sobre desempeño en juego[12]

Factor	gdl	Probabilidad	N° de elementos
Explicación	1	4,26	10,12*
Motivación	1	0,61	0,18
Expl. x Mot.	1	0,15	22,08***
Residuos	6.551	2,18	2,95

*p<0,1; **p<0,05***p<0,01; gdl: grados de libertad.

Gráfico 22: Efectos de interacción sobre Probabilidad de detección

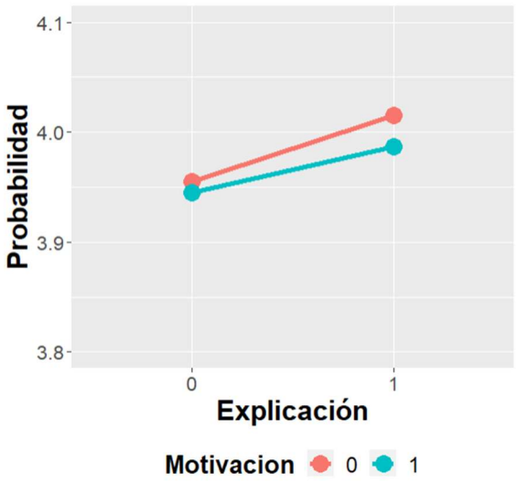
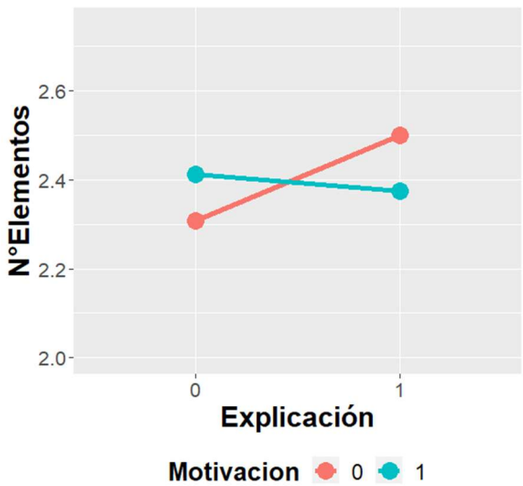


Gráfico 23: Efectos de interacción sobre N° de Elementos



Análisis gráfico Probabilidad de detección de correos fraudulentos

El **gráfico 22** representa un ANOVA factorial que analiza los efectos de *Explicación* y *Motivación* sobre la variable dependiente *Probabilidad* de detectar correos fraudulentos. En el eje X se encuentra el factor *Explicación*, con dos niveles: 0 (sin explicación) y 1 (con explicación). El eje Y muestra la *Probabilidad*, con valores más altos indicando un mejor desempeño en la tarea evaluada. Las líneas de colores representan los niveles de *Motivación*: la línea roja indica sin motivación (0) y la línea azul representa con

[12] Los valores bajo cada indicador corresponden a la suma promedio de cuadrados de los tratamientos y residuos (*mean sum of squares*, MSS), que, si se multiplica por los grados de libertad (*gdl*) resulta en la suma total de cuadrados (*total sum of squares*, TSS). Los asteriscos indican la significancia de la prueba *F*, cuyo estadístico se obtiene dividiendo la MSS correspondiente a cada tratamiento por la MSS de los residuos. De esta manera, se puede reconstruir la tabla ANOVA para cualquier indicador de resultado.

motivación (1). Ambas líneas siguen una tendencia ascendente, lo que sugiere un posible efecto positivo de la *Explicación*.

El análisis de ANOVA factorial y el análisis gráfico sugieren que ni la *Explicación* ni la *Motivación* tienen un impacto claro en la *Probabilidad.* Aunque el gráfico muestra una ligera tendencia al alza cuando se introduce la *Explicación,* esta diferencia no es estadísticamente significativa. Además, no hay evidencia de interacción entre *Explicación y Motivación*, lo que indica que estos factores no modifican la *Probabilidad* de manera sustancial en este contexto **(Tabla 21, Gráfico 22).**

Análisis gráfico N° de Elementos detectados

El **gráfico 23** representa un ANOVA factorial que analiza el impacto de los factores *Explicación y Motivación* en la variable dependiente "*Número de Elementos*". El análisis sugiere que la *Explicación* es un factor clave para mejorar el desempeño en la identificación de elementos en el juego, mientras que la *Motivación* por sí sola no tiene un impacto claro. Sin embargo, la interacción entre *Explicación y Motivación* es altamente significativa, lo que indica que su combinación puede potenciar o alterar el efecto de la *Explicación*. En particular, la *Explicación* mejora el desempeño cuando no hay *Motivación.* **(Tabla 21, Gráfico 23).**

Evaluación de las Hipótesis de Investigación sobre impactos en la Clasificación correcta de Correos Fraudulentos y Legítimos:

Los resultados obtenidos permiten rechazar las hipótesis H1 y H2, dado que la inclusión de *explicaciones* sobre técnicas de ingeniería social o *mensajes motivacionales* no generó un incremento estadísticamente significativo en la *probabilidad* de clasificar correctamente los correos fraudulentos y legítimos.

Por otro lado, los resultados no permiten rechazar la hipótesis H3, lo que indica que la presencia de e*xplicaciones* sobre ingeniería social aumenta la *cantidad de elementos identificados* por los participantes. Sin embargo, se observa una interacción significativa entre *Explicación* y *Motivación*, sugiriendo que el efecto positivo de la *Explicación* es más pronunciado en ausencia de la *Motivación.*

Finalmente, se rechaza la hipótesis H4, dado que los *mensajes motivacionales*, por sí solos, no generan un impacto claro e incluso podrían resultar contraproducentes en este contexto.

Experimento 2[13]

El análisis ANOVA muestra que el factor *juego* tiene un efecto positivo y significativo tanto en la *actitud* hacia las medidas de seguridad como en la *intención* de protegerse. Además, la interacción entre *explicación, motivación y juego* tiene un efecto positivo y significativo en la *actitud,* lo que indica que el impacto del juego se potencia cuando se combina con los otros factores. Los resultados del ANOVA confirman que el *juego* es el factor con mayor influencia en ambos indicadores ($p<0,01$). Mientras que la interacción entre los tres factores también es significativa en la actitud ($p<0,1$), no se observan efectos relevantes en la intención de protegerse **(Tabla 22 y tabla 23).**

Tabla 22. ANOVA: significancia de factores sobre evaluación de campaña

Factor	gdl	Actitud	Intención
Explicación	1	2,65	1,84
Motivación	1	0,89	1,12

[13] Los resultados del análisis de ANOVA factorial expuestos en esta sección se mantienen al aplicar el ANOVA tipo III

Juego	1	35,83***	117,48***
Expl. x Mot.	1	0,61	0
Expl. x Juego	1	0,01	0,09
Mot. x Juego	1	1,73	2,08
Expl. x Mot. x Juego	1	3,17*	0,02
Residuos	4.695	1,05	1,89

Nota: *p<0,1; **p<0,05; *** p<0,01

Tabla 23. ANOVA: significancia de factores sobre evaluación de campaña

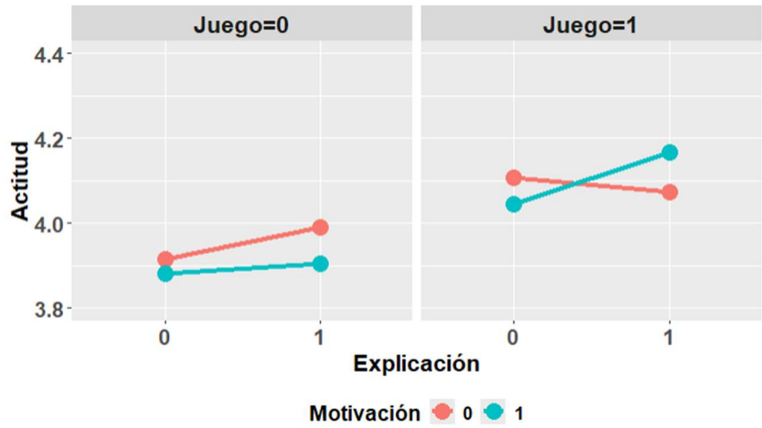
	Análisis Factorial ANOVA	
	Actitud (1)	Intención (2)
Explicación	0,075	0,028
	(0,058)	(0,078)
Motivación	-0,035	0,071
	(0,057)	(0,077)
Juego	0,191***	0,347***
	(0,060)	(0,081)
Expl. & Mot.	-0,052	0,008
	(0,082)	(0,110)
Expl. & Juego	-0,109	0,025
	(0,085)	(0,114)
Mot. & Juego	-0,026	-0,076
	(0,084)	(0,113)
Expl. & Mot. & Juego	0,208*	-0,016
	(0,120)	(0,161)
Constant	3,916***	4,012***
	(0,042)	(0,056)
N	4.703	4.703
R2	0,009	0,014
Adjusted R2	0,008	0,012
Residual Std. Error	1,023	1,375
F Statistic	6,127***	9,264***

Nota: *p<0,1; **p<0,05; *** p<0,01

A continuación, se visualizan los efectos de los 3 factores en las variables *actitud* e *intención* a través de gráficos de interacción. Respecto al **gráfico 24**, el eje X indica la presencia o ausencia de *Explicación* (0 = sin explicación, 1 = con explicación), mientras que el eje Y mide la *Actitud* de los participantes. Los dos paneles representan condiciones con *Juego* = 0 (izquierda, sin juego) y *Juego* = 1 (derecha, con juego). Además, las líneas de color diferencian la presencia de *Motivación* (rojo = sin motivación, azul = con motivación), lo que permite observar cómo interactúan estos factores en la *actitud* de los participantes.

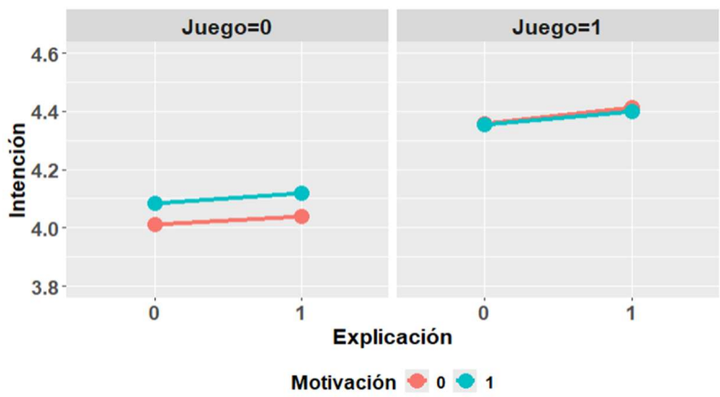
El análisis sugiere que el *Juego* es el principal factor que influye en la *Actitud*, mientras que la *Explicación* y la *Motivación* no tienen un impacto claro por sí solas. Sin embargo, la combinación de *Explicación, Motivación y Juego* puede generar cambios en la *Actitud,* aunque su efecto es menos fuerte que el del *Juego* por sí solo. Esto resalta la importancia del diseño estratégico de intervenciones, ya que la combinación de factores puede tener efectos diferenciales dependiendo del contexto. (Tabla 22, Gráfico 24).

Gráfico 24. Efectos de interacción sobre la actitud hacia la campaña



En el caso de la *intención* de protección, el análisis sugiere que el *Juego* es el principal factor que influye en la *Intención*, mientras que la *Explicación* y la *Motivación* no tienen un impacto claro por sí solas. Además, no se observan interacciones significativas entre los factores, lo que indica que el efecto del *Juego* es independiente. Estos hallazgos refuerzan la importancia del *Juego* como una herramienta clave para fomentar la *Intención* de protección, mientras que la *Explicación* y la *Motivación*, en este contexto, no generan mejoras sustanciales. **(Tabla 22, Grafico 25).**

Gráfico 25. Efectos de interacción sobre la intención de protegerse



Evaluación de las Hipótesis de Investigación sobre impactos en la Actitud hacia las medidas de seguridad e Intención de Protección:
Los resultados obtenidos confirman las hipótesis H7 y H10, lo que indica que el *Juego* es el principal factor que influye en la *Actitud e Intención*, generando mejoras significativas en ambas variables.

En contraste, se rechazan las Hipótesis asociadas a impactar la *actitud* (H5 y H6), ya que la *Explicación* y la *Motivación* no tienen un impacto claro por sí solas, lo que sugiere que su efecto individual es limitado en este contexto. Sin embargo, si hay evidencia que la interacción de los factores (*Explicación, Motivación y Juego)* mejora la *actitud* de las personas. Esto refuerza la idea que, si bien la *Explicación* y la *Motivación* pueden tener efectos individuales limitados, su eficacia depende de una integración estratégica con el *Juego.*

En el caso de la *intención*, se rechazan las hipótesis H8 y H9, ya que ni las *Explicaciones* ni los *mensajes motivacionales*, por sí solos, tienen un impacto significativo en la *Intención* de Protección. Para esta variable, el *Juego* opera de manera independiente y su efecto no se ve modificado por la *Explicación* o la *Motivación*, lo que refuerza su importancia como una herramienta clave para fomentar la protección.

En conclusión, el *Juego* es el elemento central para mejorar la Actitud e Intención de protección, mientras que la *Explicación* y la *Motivación* pueden jugar un rol complementario, particularmente en la modulación de la *Actitud.* Esto resalta la importancia del diseño estratégico de intervenciones, donde la combinación de factores puede generar efectos diferenciales dependiendo del contexto.

Variables Canal de Persuasión

En la **tabla 24** se analizan los efectos de los factores sobre los resultados intermedios. Respecto de los *efectos principales*, se encuentra que el factor *juego* actúa a través de todos los canales propuestos. La *Explicación* lo hace a través una mayor *Autoeficacia, Efectividad percibida, Gusto general por la campaña y Fuerza argumental percibida.* La *Motivación* sólo actúa a través del Gusto general, pero con un impacto negativo. Por otro lado, las interacciones de solo dos factores tienden a disminuir los impactos, en tanto, la interacción entre los tres factores aumenta la gravedad percibida y la fuerza argumental percibida. En general, se validan parcialmente los supuestos teóricos.

Tabla 24 ANOVA: significancia de factores sobre evaluación de campaña

Factor	gdl	Atención	Comprensión	Fuerza Argumental Percibida	Efectividad Percibida
Explicación	1	4,12	0,03	7,92***	7,86***
Motivación	1	0,31	0,02	0,04	0,66
Juego	1	98,1***	2,07***	11,05***	38,02***
Expl. x Mot.	1	0,28	00,85***	0,02	0,02
Expl. x Juego	1	1,9	0,00	0,17	0,35
Mot. X Juego	1	2,95	0,36*	0,03	0,06
Expl. x Mot. x Juego	1	3,03	0,18	3,49**	2,57
Residuos	4.695	1,53	0,11	0,9	0,95

Factor	gdl	Gusto General	Gravedad Percibida	Autoeficacia Percibida
Explicación	1	9,75***	0,15	9,04***
Motivación	1	3,71*	0,01	0,71
Juego	1	50,26***	3,11**	41,4***
Expl. x Mot.	1	1,13	0,41	0.00
Expl. x Juego	1	0,41	0,05	0,01
Mot. x Juego	1	0,84	0,42	1,46
Expl. x Mot. x Juego	1	0,93	2,12*	2,09
Residuos	4.695	1,22	0,67	1,08

Nota: *p<0,1; **p<0,05; *** p<0,01

Análisis Atención

Los resultados sugieren que el *Juego* es el principal determinante del aumento en Atención. La *Explicación* tiene un efecto leve, pero más notable cuando el *Juego* no está presente, mientras que la *Motivación* no genera mejoras significativas. Además, las interacciones entre factores no son significativas, lo que indica que el impacto del *Juego* es independiente de la *Explicación* y la *Motivación* **(Tabla 24, Gráfico 25).**

Análisis Comprensión

Los resultados exponen que el *Juego* es el principal factor que influye en la Comprensión. La *Explicación* y la *Motivación*, por separado, no tienen efectos significativos, pero su combinación sí genera un impacto. Además, cuando el *Juego* está presente, la *Explicación* puede incluso reducir la Comprensión en ciertos casos, dependiendo del nivel de *Motivación*. Esto sugiere que el efecto del *Juego* en la *Comprensión* es robusto, pero

las combinaciones de *Explicación y Motivación* deben considerarse con cuidado para evitar efectos contraproducentes **(Tabla 24, Gráfico 25).**

Análisis Fuerza Argumental Percibida

Los resultados sugieren que el *Juego* y la *Explicación* son los principales factores que mejoran la *Fuerza Argumental Percibida*. La *Motivación* por sí sola no tiene un efecto significativo, pero cuando se combina con *Explicación* y *Juego,* puede potenciar la *percepción de la solidez de los argumentos*. Esto resalta la importancia de integrar estos tres elementos de manera estratégica en las intervenciones para maximizar su impacto **(Tabla 24, Gráfico 25)**.

Análisis Efectividad Percibida

El análisis sugiere que el *Juego* y la *Explicación* son los principales factores que mejoran la *Efectividad Percibida*, mientras que la *Motivación* no tiene un impacto claro. Además, no se observan interacciones significativas entre los factores, lo que indica que los efectos de la *Explicación* y el *Juego* son independientes y no dependen de la *Motivación*. Estos hallazgos resaltan la importancia de incluir elementos lúdicos e informativos en campañas de sensibilización para maximizar la percepción de efectividad **(Tabla 24, Gráfico 25).**

Análisis Gusto General

El análisis de ANOVA factorial sugiere que el *Juego*, la *Explicación* y la *Motivación* tienen efectos significativos y positivos en *Gusto General*, lo que implica que estos elementos contribuyen a mejorar el *gusto general* de la campaña. Sin embargo, no se observan interacciones significativas entre los factores, lo que indica que sus efectos son independientes. Estos hallazgos resaltan la importancia de integrar estos tres elementos en estrategias de comunicación y diseño de experiencias para maximizar la aceptación y satisfacción del público **(Tabla 24, Gráfico 25).**

Análisis Gravedad Percibida

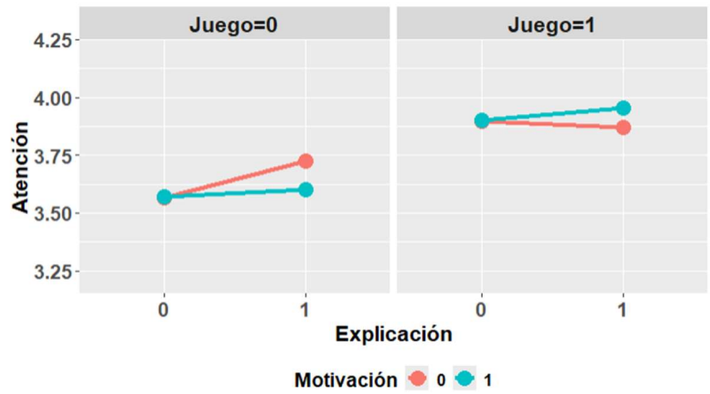
Los resultados muestran que el *Juego* tiene un impacto en la *Gravedad Percibida*, aunque de menor magnitud en comparación con otras variables. Ni la *Explicación* ni la *Motivación* tienen efectos significativos por sí solas, pero su interacción con el *Juego* sí influye en la *percepción de gravedad*. Estos resultados indican que la *percepción de gravedad* frente a los riesgos es más estable y menos sensible a los factores manipulados en este experimento, aunque ciertas combinaciones pueden modificarla en menor medida **(Tabla 24, Gráfico 25).**

Análisis Autoeficacia percibida

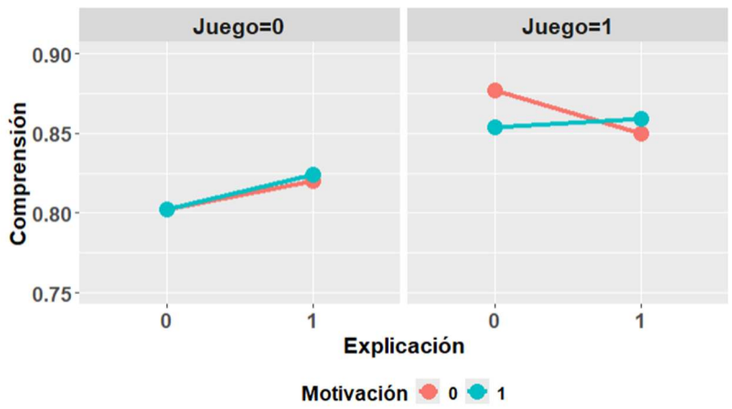
El análisis sugiere que el *Juego* y la *Explicación* son los principales factores que mejoran la *Autoeficacia Percibida*, mientras que la *Motivación* no tiene un impacto claro. Además, no se observan interacciones significativas entre los factores, lo que indica que los efectos de la *Explicación* y el *Juego* son independientes. Estos hallazgos resaltan la importancia de incluir elementos lúdicos e informativos en campañas de sensibilización para fortalecer la confianza de los participantes en su capacidad para actuar frente a riesgos **(Tabla 24, Gráfico 25).**

Gráfico 25. Efectos de interacción sobre las variables del Canal de Persuasión

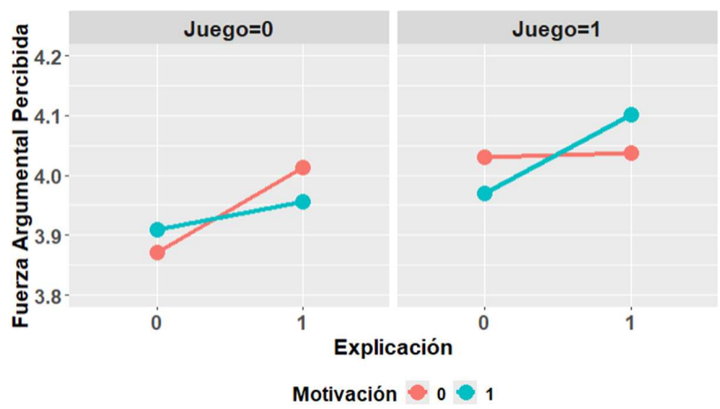
Atención



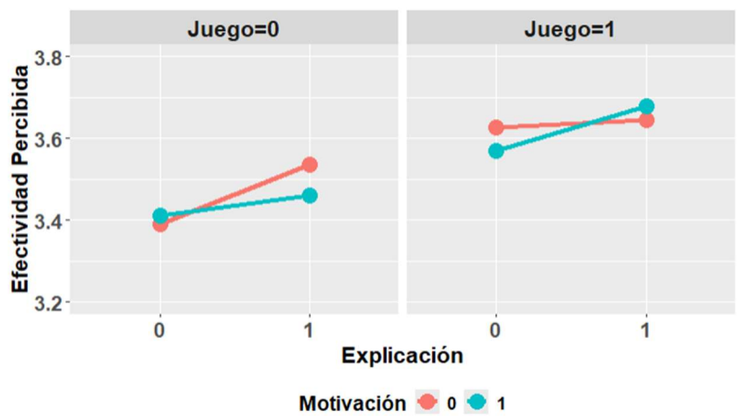
Comprensión



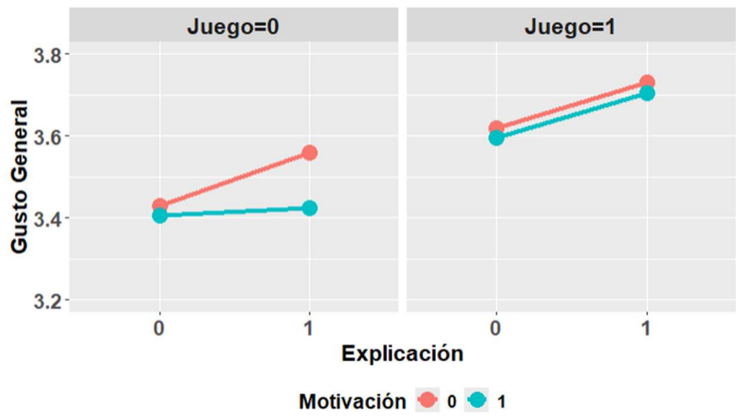
Fuerza Argumental Percibida



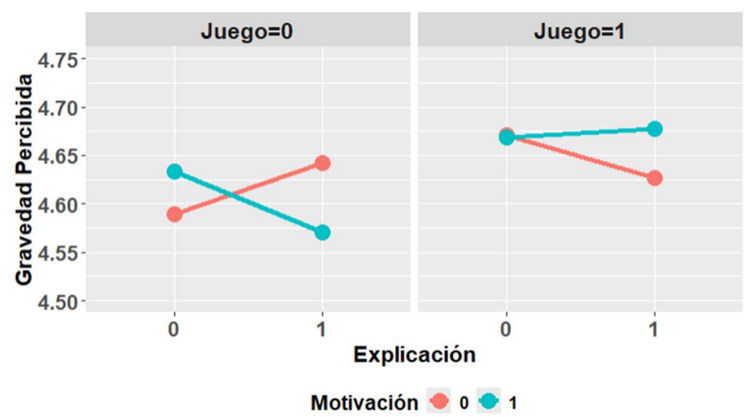
Efectividad Percibida



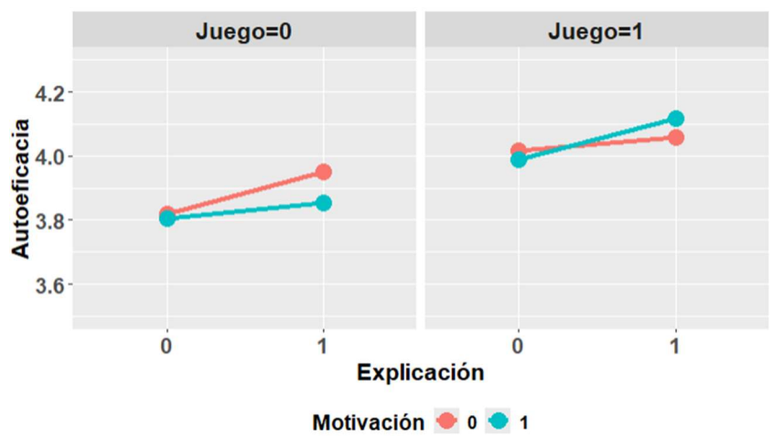
Gusto General



Gravedad Percibida



Autoeficacia Percibida



Evaluación de las Hipótesis de Investigación sobre impactos en las variables asociadas al canal de persuasión:

Se confirman parcialmente las hipótesis H11 a H15.
Los resultados confirman que el *Juego* es el factor con mayor impacto en la mayoría de las variables analizadas, mejorando significativamente *Atención, Comprensión, Fuerza Argumental Percibida, Efectividad Percibida, Gusto General y Autoeficacia Percibida.*

El factor *Explicación* también tiene un efecto positivo en varias de estas variables, especialmente cuando se combina con el Juego.

Por otro lado, la *Motivación* por sí sola no muestra un impacto significativo, aunque en combinación con el Juego y la Explicación puede potenciar algunos efectos, como en la percepción de la solidez argumentativa.

En el caso de *Gravedad Percibida*, su estabilidad sugiere que los participantes ya consideran los fraudes de phishing como un riesgo alto, por lo que los factores manipulados tienen un menor impacto en esta dimensión.

Además, la ausencia de interacciones significativas en la mayoría de los casos indica que los efectos del *Juego* y la *Explicación* operan de manera independiente. Estos hallazgos resaltan la importancia de integrar elementos lúdicos e informativos en campañas de sensibilización.

5.4 Tamaño de los Efectos

A continuación, se presenta un conjunto de indicadores que reportan el tamaño de los efectos de los tratamientos y los factores (ver sección 4.2.1.4). Sin embargo, es importante destacar que para este estudio es esperable que los tamaños de los efectos sean pequeños debido a que las variables dependientes miden percepciones mediante escalas de Likert de 5 puntos, lo que introduce limitaciones de medición. Estas escalas restringen la capacidad de respuesta a pocos niveles discretos, reduciendo la sensibilidad a cambios sutiles en actitudes o percepciones, lo que puede generar una subestimación de los efectos reales. Russell y Bobko (1992) encontraron que las escalas de 5 puntos eran rígidas para captar efectos moderadores sutiles y que, al aumentar la granularidad, los tamaños del efecto crecían significativamente. Así, la combinación de una escala ordinal de baja resolución, su poca sensibilidad a cambios sutiles y la varianza restringida, junto con la naturaleza moderada de muchos fenómenos psicológicos, puede explicar la tendencia a obtener tamaños de efecto reducidos en este tipo de medición.

Modelos basados en Tratamientos

Experimento 1: En general, los tratamientos no parecen tener un efecto significativo sobre la variable probabilidad y en el caso del número de elementos identificados, sólo Motivación (T2) y Explicación (T3) tienen un efecto significativo en el número de elementos, cuyos tamaños de efecto son 0,06 y 0,11 respectivamente. De acuerdo con las directrices de Cohen, estos valores se consideran muy pequeños, ya que un tamaño de efecto de 0,2 se clasifica como pequeño, 0,5 como mediano y 0,8 como grande **(Tabla 25).**

Tabla 25: Tamaños del efecto (Cohen's d) en clasificación de correos

Tratamiento	Explicación	Motivación	Probabilidad (Cohen's d)	N° de elementos (Cohen's d)
T2	0	1	-0,01	0,06*
T3	1	0	0,04	0,11***
T4	1	1	0,02	0,04

Nota: *p<0,1; **p<0,05; *** p<0,01

Experimento 2: Los resultados muestran que la presencia de un juego en la intervención es clave para mejorar tanto la actitud como la intención de los participantes, ya que todas las condiciones con juego presentan efectos positivos y significativos, mientras que las condiciones sin juego no generan impacto relevante. La combinación de explicación + motivación + juego (T4) obtiene el mayor efecto en actitud (d = 0,25), sugiriendo que estos elementos pueden potenciarse entre sí. En contraste, las condiciones sin juego presentan efectos mínimos o nulos, lo que indica que el juego es un componente esencial para influir en el comportamiento de los participantes **(Tabla 26).**

Tabla 26: Tamaños del efecto (Cohen's d) de Impactos en indicadores de Actitud e Intención

Tratamiento	Explica-ción	Motiva-ción	Juego	Actitud (Cohen's d)	Intención (Cohen's d)
T1	0	0	1	0,19***	0,25***
T2	0	1	1	0,12**	0,24***
T3	1	0	1	0,15***	0,29***
T4	1	1	1	0,25***	0,28***
T6	0	1	0	-0,03	0,05
T7	1	0	0	0,07	0,02
T8	1	1	0	-0,01	0,07

Nota: *p<0,1; **p<0,05; *** p<0,01

Respecto a los resultados asociados a las variables del canal de persuasión, los resultados presentados en la **tabla 27** muestran que T4 presenta los efectos más altos en atención (d = 0,31), fuerza argumental percibida (d = 0,24), efectividad percibida (d = 0,29), gravedad percibida (d = 0,11), y autoeficacia percibida (d = 0,29). En tanto, T3 logra un mejor desempeño en Gusto General (d = 0,28) y T1 logra un mayor efecto en Comprensión (d = 0,27). De acuerdo con los criterios de Cohen, estos efectos se consideran pequeños. En contraste, T6 y T8 presentan valores cercanos a cero en la mayoría de las variables, sin efectos significativos, lo que indica que no generan impacto relevante.

Tabla 27 Tabla 26: Tamaños del efecto (Cohen's d) de Impactos en indicadores del Canal de persuasión

Condición Experimental	Atención	Comprensión	Fuerza Argumental Percibida	Efectividad Percibida	Gusto General	Gravedad Percibida	Autoeficacia Percibida
T1	0,27***	0,27***	0,17***	0,24***	0,17***	0,10*	0,20***
T2	0,27***	0,18***	0,10*	0,18***	0,15***	0,10	0,16***
T3	0,25***	0,17***	0,18***	0,26***	0,28***	0,04	0,23***
T4	0,31***	0,20***	0,24***	0,29***	0,25***	0,11*	0,29***
T6	0,00	0,00	0,04	0,02	-0,02	0,05	0,01
T7	0,12**	0,06	0,15**	0,15***	0,12**	0,06	0,12**
T8	0,03	0,07	0,09	0,07	0,00	-0,02	0,03

Nota: *p<0,1; **p<0,05; *** p<0,01

Modelo basado en Factores

Experimento 1: Los resultados de la **Tabla 28** muestran los valores de Eta cuadrado parcial (η^2 parcial), que representan el tamaño del efecto de los factores *Explicación, Motivación* y su interacción (*Explicación* × *Motivación*) sobre las variables *Probabilidad y Número de Elementos.* Los resultados indican que ninguno de los factores analizados tiene un impacto significativo en la Probabilidad ni en el Número de Elementos, ya que los tamaños de efecto son extremadamente pequeños, lo que sugiere una influencia marginal en ambas variables.

Tabla 28. Tamaño del efecto (Eta2 parcial)

Factor	Probabilidad	N° de elementos
Explicación	0,0003	0,0005
Motivación	0,0000	0,0000
Expl. x Mot.	0,0000	0,0011

Experimento 2: Los resultados de la **Tabla 29** presentan los valores de Eta cuadrado parcial (η² parcial), que indican el tamaño del efecto de los factores *Explicación, Motivación, Juego* y sus interacciones sobre las variables *Actitud e Intención*. Los resultados sugieren que *Juego* es el único factor con un impacto perceptible en *Actitud e Intención*, aunque sigue siendo pequeño en términos de magnitud de efecto. *Explicación, Motivación* y sus combinaciones no tienen una influencia significativa en ninguna de las variables, lo que indica que estos factores por sí solos o en combinación no generan cambios sustanciales en la *actitud o intención* de los participantes.

Tabla 29. Tamaño del efecto (Eta2 parcial)

Factor	Actitud	Intención
Explicación	0,0005	0,0002
Motivación	0,0001	0,0001
Juego	0,0072	0,0131
Expl. x Mot.	0,0002	0,0000
Expl. x Juego	0,0000	0,0000
Mot. x Juego	0,0004	0,0002
Expl. x Mot. x Juego	0,0006	0,0000

Los resultados de la **Tabla 30** presentan los valores de Eta cuadrado parcial (η² parcial), que indican el tamaño del efecto de los factores *Explicación, Motivación, Juego* y sus interacciones sobre diversas variables dependientes: *Atención, Comprensión, Fuerza Argumental Percibida, Efectividad Percibida, Gusto General, Gravedad Percibida y Autoeficacia.*

Los resultados sugieren que *Juego* es el único factor con un efecto perceptible, aunque pequeño (η_p^2=0,01), en *Atención, Efectividad Percibida, Gusto General y Autoeficacia*. *Explicación* tiene un impacto mínimo en *Fuerza Argumental Percibida y Autoeficacia*, mientras que *Motivación* no genera cambios relevantes. Las interacciones entre factores presentan tamaños de efecto muy bajos, lo que indica que las combinaciones de *Explicación, Motivación y Juego* no generan un impacto sustancial en las percepciones de los participantes.

Tabla 30. Tamaño del efecto (Eta2 parcial)

Factor	Atencion	Compren-sión	Fuerza Argumental Percibida	Efectivi-dad Perc.	Gusto General	Gravedad Percibida	Autoefi-cacia
Explica-ción	0,0005	0,0001	0,0018	0,0017	0,0017	0,0000	0,0018
Motiva-ción	0,0000	0,0000	0,0000	0,0001	0,0005	0,0000	0,0001
Juego	0,0135	0,0075	0,0026	0,0085	0,0088	0,0010	0,0081
Expl. x Mot.	0,0000	0,0003	0,0000	0,0000	0,0002	0,0001	0,0000
Expl. x Juego	0,0003	0,0008	0,0000	0,0001	0,0001	0,0000	0,0000
Mot. x Juego	0,0004	0,0001	0,0000	0,0000	0,0001	0,0001	0,0003
Expl. x Mot. x Juego	0,0004	0,0001	0,0008	0,0006	0,0002	0,0007	0,0004

Modelo no Paramétrico

La **tabla 31** presenta los coeficientes de r de Rank-Biserial, que indican el tamaño del efecto. A la vez, se presentan los resultados de la Prueba Post Hoc Dunn (significancia

estadística) junto con sus p-valores entre paréntesis. Se comparan diferentes grupos en variables como Actitud, Intención, Atención, Comprensión, Fuerza Argumental Percibida, Efectividad Percibida, Gusto General, Gravedad Percibida y Autoeficacia Percibida.

Los resultados de la Prueba Post Hoc Dunn muestran diferencias significativas en varias variables, *siendo Intención, Atención, Efectividad Percibida y Autoeficacia Percibida* las más afectadas, con p-valores muy bajos en múltiples comparaciones (especialmente en 1-5, 3-5 y 4-5). *Fuerza Argumental Percibida y Comprensión* presentan algunas diferencias, aunque con efectos más pequeños o marginalmente significativos. *Gusto General y Actitud* muestran efectos moderados, mientras que *Gravedad Percibida* no tiene diferencias significativas en ninguna comparación. En general, los tratamientos analizados parecen tener un impacto diferencial en las percepciones y actitudes de los participantes, con efectos más marcados en variables relacionadas con intención y percepción de efectividad.

La interpretación de r sigue los criterios estándar de magnitud: 0,1 pequeño, 0,3 mediano, 0,5 grande. Esto permite comparar la fuerza de las diferencias entre distintas parejas de grupos.

Tabla 31. Tamaño del efecto (Prueba Post Hoc Dunn)

	1-5 r de Rank-Biserial	2-5 r de Rank-Biserial	3-5 r de Rank-Biserial	4-5 r de Rank-Biserial	6-5 r de Rank-Biserial	7-5 r de Rank-Biserial	8-5 r de Rank-Biserial
Actitud	0,052** (0,011)	0,043* (0,083)	0,037 (0,298)	0,063*** (0,000)	0,011 (1,000)	0,017 (1,000)	0,003 (1,000)
Intención	0,080*** (0,000)	0,082*** (0,000)	0,093*** (0,000)	0,086*** (0,000)	0,008 (1,000)	0,005 (1,000)	0,019 (1,000)
Atención	0,060*** (0,001)	0,068*** (0,000)	0,059*** (0,001)	0,080*** (0,000)	0,000 (1,000)	0,032 (0,765)	0,006 (1,000)
Comprensión	0,066*** (0,000)	0,044* (0,073)	0,041 (0,133)	0,047** (0,037)	0,006 (1,000)	0,015 (1,000)	0,021 (1,000)
Fuerza Argumental Percibido	0,039 (0,205)	0,034 (0,541)	0,043* (0,087)	0,063*** (0,000)	0,009 (1,000)	0,044* (0,075)	0,023 (1,000)
Efectividad Percibida	0,063*** (0,000)	0,053*** (0,009)	0,067*** (0,000)	0,078*** (0,000)	0,004 (1,000)	0,040 (0,179)	0,019 (1,000)
Gusto General	0,046** (0,047)	0,043* (0,087)	0,070*** (0,000)	0,066*** (0,000)	0,003 (1,000)	0,032 (0,796)	0,001 (1,000)
Gravedad Percibida	0,027 (1,000)	0,025 (1,000)	0,012 (1,000)	0,030 (1,000)	0,012 (1,000)	0,012 (1,000)	0,011 (1,000)
Autoeficacia Percibida	0,043* (0,083)	0,049** (0,024)	0,059*** (0,001)	0,079*** (0,000)	0,000 (1,000)	0,040 (0,168)	0,009 (1,000)

Nota: *p<0,1; **p<0,05; *** p<0,01

6. Principales Conclusiones

En la actualidad, el fraude financiero cibernético representa una amenaza global de gran envergadura, que afecta a individuos, empresas y gobiernos. Según el *World Economic Forum*, la ciberseguridad figura entre los riesgos globales más críticos a corto plazo (WEF, 2024). Entre los fraudes más frecuentes destaca el *phishing*, una modalidad que busca manipular a las víctimas para que realicen acciones específicas —como revelar información financiera— mediante avanzadas técnicas de ingeniería social y subterfugios técnicos.

En este contexto, la reciente Ley N° 21.673, publicada en el Diario Oficial el 30 de mayo de 2024, reformuló el marco regulatorio relativo a la limitación de la responsabilidad de los usuarios de medios de pago frente al fraude, establecido en la Ley N° 20.009. Esta nueva normativa introdujo, entre otros aspectos, responsabilidades tanto para los usuarios como para las entidades reguladas. Por una parte, la reforma dispone que "los usuarios deberán informarse y adoptar todas las medidas necesarias para prevenir el uso indebido, el fraude u otros riesgos afines a la utilización de los medios de pago a que se refiere esta ley y los mecanismos de autenticación asociados". Paralelamente, exige que las entidades reguladas proporcionen información periódica, clara, accesible y actualizada sobre las medidas de seguridad y las instrucciones para un uso seguro, fomentando prácticas responsables en la gestión de los medios de pago (art. 4 bis).

A nivel local, muchas instituciones financieras no ofrecen campañas específicas sobre fraude en medios de pago en línea o, si las tienen, suelen ser parciales, no abordan todos los tipos de fraudes a los que están expuestos los usuarios y son difíciles de acceder (SERNAC, 2025a). Tampoco se evalúa su efectividad, lo cual resulta problemático si consideramos que, según extractos de reclamos ingresados al SERNAC, los usuarios pueden llegar a interpretar de manera restrictiva y errónea las recomendaciones entregadas (SERNAC, 2025a).

En consecuencia, el presente estudio tuvo como objetivo identificar y evaluar elementos comunicacionales que pueden optimizar la efectividad de las campañas contra el phishing. Se analiza cómo estos factores contribuyen, por un lado, a fortalecer la capacidad de las personas para clasificar correctamente los correos que reciben (entre fraudulentos y legítimos) y, por otro, a influir positivamente en sus actitudes y percepciones sobre la seguridad digital. Todo ello con el propósito de fomentar una mayor disposición a adoptar, de manera preventiva, las recomendaciones de seguridad.

Para este fin, se llevó a cabo un **Experimento Controlado Aleatorizado** (RCT, por sus siglas en inglés), un diseño que facilita el establecimiento de relaciones causales entre las variables de interés. Bajo este marco metodológico, se realizaron evaluaciones de impacto a nivel de tratamiento (*Cell Means Model*) y a nivel de factores (ANOVA factorial).

En concreto, se evaluó el impacto de tres estrategias comunicacionales diseñadas para reforzar una campaña estándar de detección de correos fraudulentos:

a) ***Mensajes motivacionales:*** Se trata de mensajes que refuerzan la autoeficacia y destacan la gravedad del *phishing*. Un estudio previo de SERNAC analizó qué factores motivan a los consumidores a protegerse contra el fraude informático, tomando como base la Teoría de Motivación de Protección (TMP). Dicho estudio concluyó que la autoeficacia es el factor más determinante para fomentar la autoprotección, seguida de la percepción de la gravedad del fraude (SERNAC, 2025b).

b) ***Explicación de técnicas de ingeniería social***: Se ofrece una breve descripción de los métodos de manipulación empleados en correos fraudulentos, lo que facilitaría su identificación.

c) ***Enfoque de juego***: Consiste en aplicar un test para evaluar la legitimidad de los correos, permitiendo a los participantes poner en práctica las recomendaciones brindadas y, a la vez, favorecer un proceso de persuasión comunicacional más efectivo, además de potenciar la inmersión en la narrativa de la campaña.

Estas estrategias comunicacionales buscan fomentar una mayor disposición de las personas a adoptar de forma preventiva las recomendaciones de seguridad. Para ello, se midieron las variables "***actitud hacia las medidas de detección de estafas en línea" e "intención de protegerse en el futuro",*** ambas consideradas precursoras del comportamiento observado. Adicionalmente, se analizaron variables intermedias relacionadas con el proceso de persuasión (atención, comprensión, aceptación general hacia la campaña, autoeficacia y gravedad percibida), lo que permitió reforzar las conclusiones sobre la efectividad persuasiva de las intervenciones.

Los hallazgos, a nivel de factores, evidenciaron que **el "Juego" constituye el componente más influyente para mejorar los indicadores de *actitud* e *intención*. así como las variables intermedias ligadas al canal de persuasión.** Su impacto es consistente y estadísticamente significativo, destacando su potencial como herramienta fundamental para reforzar la percepción y la conducta en ámbitos de ciberseguridad y protección contra fraudes.

Por otra parte, **la "Explicación" también cumple una función relevante al elevar la aceptación general de la campaña y la percepción de autoeficacia; no** obstante, su influencia se torna más notable cuando el "Juego" está ausente o al combinarse con la "Motivación".

Finalmente, **la "Motivación"** propuesta en este estudio **por si sola no arrojó resultados con significancia estadística en la mayoría de las variables**, lo que sugiere que, si bien los mensajes motivacionales pueden realzar la percepción de gravedad y la autoeficacia, **no son suficientes para inducir cambios trascendentes en la actitud o en el comportamiento preventivo.** Sin embargo, al combinarse con la "Explicación", se aprecian mejoras sustanciales en su impacto, lo que indica que **la efectividad de la "Motivación" depende de su adecuada integración con otros elementos informativos.**

Los análisis a nivel de tratamiento evidenciaron que todos los tratamientos que incorporaron el factor *Juego* generaron efectos positivos sobre la *actitud e intención de protección*, resultando estadísticamente significativos ($p < 0,01$). En contraste, los tratamientos que no incluyeron este factor no mostraron significancia estadística en estas variables. Asimismo, estos resultados fueron respaldados por pruebas de robustez tanto paramétricas como no paramétricas.

En cuanto a los análisis de heterogeneidad, los resultados indican que los impactos de los tratamientos en las variables de *actitud e intención* no presentan diferencias según el género; dicho de otro modo, **los efectos sobre la actitud hacia las medidas de seguridad y la intención de protegerse son independientes del género de los participantes.**

Respecto a la heterogeneidad por edad (menores o mayores de 45 años), los hallazgos sugieren que, **aunque los adultos jóvenes muestran una disposición inicial más baja hacia la seguridad, las estrategias que combinan juego con motivación y/o explicación pueden mejorar tanto su actitud como su intención de protegerse.** En cuanto a las distintas estimaciones del tamaño de los efectos, tanto a nivel de tratamiento como a nivel de factores, se concluye que los resultados obtenidos fueron de magnitud pequeña. Sin embargo, en este estudio era previsible encontrar efectos de esta naturaleza, dado que las variables dependientes se miden a través de escalas de Likert de 5 puntos, lo que introduce limitaciones de medición. Dichas escalas restringen las respuestas a un número reducido de niveles discretos, disminuyendo la sensibilidad a cambios sutiles en actitudes o percepciones y, por tanto, pueden subestimar los efectos reales.

Finalmente, respecto a las **recomendaciones de política** que se pueden derivar de este estudio, se propone lo siguiente:

1. Reforzar las campañas de concientización contra el fraude online mediante nuevos recursos comunicacionales que fomenten la adopción preventiva de medidas de seguridad.

En particular, se recomienda el uso de juegos serios en sus diversas modalidades, ya que incorporan mecanismos clave que favorecen el aprendizaje y la transformación de actitudes y comportamientos.

Asimismo, en las campañas dirigidas contra el phishing —caracterizado por el empleo de técnicas de manipulación sofisticadas— conviene explicar detalladamente las tácticas de ingeniería social más habituales, de modo que las personas reconozcan la manipulación y actúen con mayor conciencia.

Por último, se sugiere incluir frases motivacionales que destaquen la autoeficacia y la gravedad del phishing, como un complemento a otros elementos comunicacionales.

2. Utilizar diversos medios para difundir la información —correo electrónico, videos, sitio web de SERNAC y portales de instituciones financieras— y personalizar la entrega de mensajes conforme al perfil sociodemográfico de cada consumidor.

3. Evaluar periódicamente las campañas a través de juegos interactivos y encuestas de percepción, con el fin de asegurar estándares mínimos y realizar los ajustes necesarios cuando los resultados no cumplan con lo esperado.

7. Bibliografía

Abt, C. C. (1974). Serious games. Viking Press.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (1975). Belief, attitude, intention and behaviour: An introduction to theory and research (Vol. 27). Addison-Wesley.

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. En The handbook of attitudes (pp. 173–221). Lawrence Erlbaum Associates Publishers.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. Journal of Experimental Social Psychology, 22(5), 453–474. https://doi.org/10.1016/0022-1031(86)90045-4

Albladi, S., & Weir, G. R. S. (2018). User characteristics that influence judgment of social engineering attacks in social networks. Human-centric Computing and Information Sciences, 8(1), 5. https://doi.org/10.1186/s13673-018-0128-7

Alessi, S. M., & Trollip, S. R. (2001). Multimedia for learning: Methods and development (3rd ed.). Allyn & Bacon. http://catalog.hathitrust.org/api/volumes/oclc/44676102.html

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 Access, 10, 1-1. https://doi.org/10.1109/ACCESS.2022.3230286.

Anderson, C. L., & Agarwal, R. (2010). Practicing safe computing: A multimedia empirical examination of home computer user security behavioral intentions. MIS Quarterly, 34(3), 613–643.

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press. https://doi.org/10.2307/j.ctvcm4j72

APWG. (2024). Phishing Activity Trends Reports: 3rd Quarter 2024. Anti-Phishing Working Group, Inc. https://apwg.org/trendsreports/

Bada, M., Sasse, A. M., & Nurse, J. R. C. (2019). Cyber Security Awareness Campaigns: Why do they fail to change behaviour? (arXiv:1901.02672). arXiv. https://doi.org/10.48550/arXiv.1901.02672

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory (pp. xiii, 617). Prentice-Hall.

Batzos, Z., Saoulidis, T., Margounakis, D., Fountoukidis, E., Grigoriou, E., Moukoulis, A., Sarigiannidis, A., Liatifis, A., Karypidis, P.-A., Bibi, S., Filippidis, A., Kazanidis, I., Nifakos, S., Kasig, T., Heydari, M., & Mouratidis, H. (2023). Gamification and Serious Games for Cybersecurity Awareness and First Responders Training: An overview. https://doi.org/10.36227/techrxiv.22650952.v1

Bitrián, P., Buil, I., Catalán, S., & Hatfield, S. (2023). The use of gamification strategies to enhance employees' attitudes towards e-training systems. The International Journal of Management Education, 21(3), 100892. https://doi.org/10.1016/j.ijme.2023.100892.

Bitrián, P., Buil, I., Catalán, S., & Merli, D. (2024). Gamification in workforce training: Improving employees' self-efficacy and information security and data protection behaviours. Journal of Business Research, 179, 114685. https://doi.org/10.1016/j.jbusres.2024.114685.

Blanco Abarca, A., Horcajo Rosado, F. J., & Sánchez Fernández, F. (2017). Cognición social. https://dialnet.unirioja.es/servlet/libro?codigo=674610

Boncu, Ștefan, Candel, O.-S., & Popa, N. L. (2022). Gameful Green: A Systematic Review on the Use of Serious Computer Games and Gamified Mobile Apps to Foster Pro-Environmental Information, Attitudes and Behaviors. Sustainability, 14(16), Article 16. https://doi.org/10.3390/su141610400.

Brehm, J. W. (1966). A theory of psychological reactance (pp. x, 135). Academic Press.

Brilingaitė, A., Bukauskas, L., Domarkienė, I., Rančelis, T., Ambrozaitytė, L., Pirta, R., Lugo, R. G., & Knox, B. J. (2025). Towards projection of the individualised risk assessment for the cybersecurity workforce. Computer Standards & Interfaces, 93, 103962. https://doi.org/10.1016/j.csi.2024.103962

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. Interactive Learning Environments, 24(6), 1162-1175. https://doi.org/10.1080/10494820.2014.964263.

Bueno, C., & Osuna, J. (2013). Evaluación del diseño de políticas públicas: propuesta de un modelo integral. Revista del CLAD Reforma y Democracia, 57, 37–66. https://www.redalyc.org/articulo.oa?id=357533689002

Buller, D. B., Borland, R., & Burgoon, M. (1998). Impact of behavioral intention on effectiveness of message features: Evidence from the Family Sun Safety Project. Human Communication Research, 24(3), 433–453. https://doi.org/10.1111/j.1468-2958.1998.tb00424.x

Burkey, D., Anastasio, D., & Suresh, A. (2013). Improving student attitudes toward the capstone laboratory course using gamification. 2013 ASEE Annual Conference & Exposition Proceedings, 23.718.1–23.718.18. https://peer.asee.org/improving-student-attitudes-toward-the-capstone-laboratory-course-using-gamification.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. Journal of Personality and Social Psychology, 42(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. Psychological Bulletin, 119, 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Cady, F. (2017). The data science handbook. Wiley.

Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? Survey Practice, 3(6). https://doi.org/10.29115/SP-2010-0028

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: Methods and applications. Cambridge University Press.

Campbell, J. P., & Kuncel, N. R. (2002). Individual and team training. En N. Anderson & D. S. Ones (Eds.), Handbook of industrial, work and organizational psychology (pp. 278–312). SAGE.

Caputo, D. D., Pfleeger, S., Freeman, J. D., & Johnson, M. (2014). Going spear phishing: Exploring embedded training and awareness. IEEE Security & Privacy, 12(1), 28–38. https://doi.org/10.1109/MSP.2013.106

Carcioppolo, N., Wendorf, J., & Tran, L. (2015). Serious Games, Health, and Organizing. En Organizations, Communication, and Health. Routledge.

Charsky, D. (2010). From edutainment to serious games: A change in the use of game characteristics. Games and Culture, 5(2), 177–198. https://doi.org/10.1177/1555412009354727

Cialdini, R. B. (2007). Influence: The psychology of persuasion (Revised ed.). Harper Business.

Cialdini, R. B. (2009). Influence: Science and practice (5th ed.). Pearson/Allyn & Bacon.

Clement, S. L., Severin-Nielsen, M. K., & Shamshiri-Petersen, D. (2020). Device effects on survey response quality: A comparison of smartphone, tablet and PC responses on a cross-sectional probability sample. Departamento de Política y Sociedad, Universidad de Aalborg, Dinamarca

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. Mass Communication and Society, 4(3), 245–264. https://doi.org/10.1207/S15327825MCS0403_01

DAS, S. (2017). Social Cybersecurity: Reshaping Security Through An Empirical Understanding of Human Social Behavior [Tesis doctoral, Carnegie Mellon University]. https://doi.org/10.1184/R1/6722918.v1

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. Plenum.

Deci, E. L., Ryan, R. M., Gagné, M., Leone, D. R., Usunov, J., & Kornazheva, B. P. (2001). Need satisfaction, motivation, and well-being in the work organizations of a former Eastern Bloc country. Personality and Social Psychology Bulletin, 27, 930–942.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." En Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (pp. 9–15). ACM.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. Journal of Personality and Social Psychology, 86(2), 345–355. https://doi.org/10.1037/0022-3514.86.2.345

Duflo, E., Glennerster, R., & Kremer, M. (2006). Using Randomization in Development Economics Research: A Toolkit. The Abdul Latif Jameel Poverty Action Lab. https://www.povertyactionlab.org/sites/default/files/research-paper/Using-Randomization-in-Development-Economics.pdf

Dunn, O. J. (1964). Multiple comparisons using rank sums. Technometrics, 6(3), 241–252. https://doi.org/10.1080/00401706.1964.10490181

Eagly, A. H., & Chaiken, S. (1993). The psychology of attitudes (pp. xxii, 794). Harcourt Brace Jovanovich College Publishers.

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. En The handbook of social psychology (Vols. 1–2, 4.ª ed., pp. 269–322). McGraw-Hill.

Edwards, J. (2024). Mastering cybersecurity: Strategies, technologies, and best practices. Apress. https://doi.org/10.1007/979-8-8688-0297-3

Egashira, M., Son, D., & Ema, A. (2022). Serious game for change in behavioral intention toward lifestyle-related diseases: Experimental study with structural equation modeling using the theory of planned behavior. JMIR Serious Games, 10(1), e28982. https://doi.org/10.2196/28982

Ezezika, O., Oh, J., Edeagu, N., & Boyo, W. (2018). Gamification of nutrition: A preliminary study on the impact of gamification on nutrition knowledge, attitude, and behaviour of adolescents in Nigeria. Nutrition and Health, 24(3), 137–144. https://doi.org/10.1177/0260106018782211

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. En M. P. Zanna (Ed.), Advances in Experimental Social Psychology (Vol. 23, pp. 75–109). Academic Press. https://doi.org/10.1016/S0065-2601(08)60318-4

Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. En Dual-process theories of the social mind (pp. 155–171). The Guilford Press.

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. En L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 14, pp. 161–202). Academic Press. https://doi.org/10.1016/S0065-2601(08)60372-X

Ferreira, A., Coventry, L., & Lenzini, G. (2015). Principles of Persuasion in Social Engineering and Their Use in Phishing.

Field, A. (2018). Discovering statistics using SPSS (5.ª ed.). SAGE Publications.

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behaviour: An introduction to theory and research (Vol. 27). Addison-Wesley.

Fishbein, M., & Ajzen, I. (1982). Acceptance, yielding and impact: Cognitive processes in persuasion. En Cognitive responses in persuasion. Psychology Press.

Flavell, J. H., Botkin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. (1968). The development of role-taking and communication skills in children (pp. vii, 239). John Wiley & Sons.

Fonfría Mesa, A., & Duch Brown, N. (2020). Ciberseguridad económica. Análisis del Real Instituto Elcano (ARI), 105, 1.

Furnell, S. M., Jusoh, A., & Katsabas, D. (2006). The challenges of understanding and using security: A survey of end-users. Computers & Security, 25(1), 27–35. https://doi.org/10.1016/j.cose.2005.12.004

Gee, J., & Gee, J. P. (2007). Social linguistics and literacies: Ideology in discourses (3.ª ed.). Routledge. https://doi.org/10.4324/9780203944806

Gerrig, R. J. (1993). Experiencing narrative worlds: On the psychological activities of reading (pp. xi, 273). Yale University Press.

Gertler, P., Martinez, S., Rawlings, L. B., Premand, P., & Vermeersch, C. M. J. (2016). Impact evaluation in practice: Second edition. IDB Publications. https://doi.org/10.18235/0006529

Gilbert, D. T. (1991). How mental systems believe. American Psychologist, 46(2), 107–119. https://doi.org/10.1037/0003-066X.46.2.107

Glennerster, R., & Takavarasha, K. (2013). Running randomized evaluations: A practical guide. Princeton University Press. https://doi.org/10.2307/j.ctt4cgd52

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. American Psychologist, 54(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. Journal of Personality and Social Psychology, 73(1), 186–199. https://doi.org/10.1037/0022-3514.73.1.186

Gragg, D. (2003). A multi-level defense against social engineering. SANS Institute - InfoSec Reading Room.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. Journal of Personality and Social Psychology, 79(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. En Psychological foundations of attitudes (pp. 147–170). Elsevier. https://doi.org/10.1016/B978-1-4832-3071-9.50012-X

Grevelink, J. (2015). Serious games for cybersecurity. https://arno.uvt.nl/show.cgi?fid=136774

Grobler, M., Gaire, R., & Nepal, S. (2021). User, usage and usability: Redefining human centric cyber security. Frontiers in Big Data, 4. https://doi.org/10.3389/fdata.2021.583723

Hagtvedt, H., & Brasel, S. A. (2017). Color saturation increases perceived product size. Journal of Consumer Research, ucx039. https://doi.org/10.1093/jcr/ucx039

Hammady, R., & Arnab, S. (2022). Serious gaming for behaviour change: A systematic review. Information, 13(3), Article 3. https://doi.org/10.3390/info13030142

Harrison, V., Kemp, R., Brace, N., & Snelgar, R. (2020). SPSS for psychologists (7.ª ed.). Bloomsbury Publishing.

Henderson, N., Pallett, H., Linden, S., Montanarini, J., & Buckley, O. (2024). The disPHISHinformation game: Creating a serious game to fight phishing using blended design approaches. En Human factors in cybersecurity (Vol. 127, pp. 146–156). https://openaccess.cms-conferences.org/publications/book/978-1-964867-03-8/article/978-1-964867-03-8_14

Herzing, J. M. E. (2019). Mobile web surveys. FORS Guide Nº 01, Versión 1.0. https://doi.org/10.24449/FG-2019-00001

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). Cochrane handbook for systematic reviews of interventions (Versión 6.0). Cochrane.

Hong, Y., & Furnell, S. (2021). Understanding cybersecurity behavioral habits: Insights from situational support. Journal of Information Security and Applications, 57, 102710. https://doi.org/10.1016/j.jisa.2020.102710

Hovland, C. I., & Janis, I. L. (1959). Personality and persuasibility (pp. xiv, 333). Yale University Press.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). Communication and persuasion: Psychological studies of opinion change (pp. xii, 315). Yale University Press.

Jari, M. (2022). An overview of phishing victimization: Human factors, training and the role of emotions. Computer Science and Information Technology, 12(13), 217–228. https://doi.org/10.5121/csit.2022.121319

Kelley, K., & Preacher, K. J. (2012). On effect size. Psychological Methods, 17(2), 137–152. https://doi.org/10.1037/a0028086

Kelley, R. L., Osborne, W. J., & Hendrick, C. (1974). Role-taking and role-playing in human communication. Human Communication Research, 1(1), 62–74. https://doi.org/10.1111/j.1468-2958.1974.tb00254.x

Khan, N. F., Ikram, N., Murtaza, H., & Javed, M. (2023). Evaluating protection motivation based cybersecurity awareness training on Kirkpatrick's model. Computers & Security, 125, 103049. https://doi.org/10.1016/j.cose.2022.103049

Knowles, E. S., & Linn, J. A. (2004). The promise and future of resistance and persuasion. En Resistance and persuasion (pp. 301–310). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410609816

Kraska-Miller, M. (2013). Nonparametric statistics for social and behavioral sciences. CRC Press. https://doi.org/10.1201/b16188

Larson-Hall, J. (2015). A guide to doing statistical analysis in second language research using SPSS (2.ª ed.). Routledge. https://doi.org/10.4324/9781315775661

Langsrud, Ø. (2003). ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. Statistics and Computing, 13(2), 163–167. https://doi.org/10.1023/A:1023260610025

Lawson, J. (2014). Design and analysis of experiments with R (2.ª ed.). CRC Press.

Liang, H., & Xue, Y. (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective. Journal of the Association for Information Systems, 11(7). https://doi.org/10.17705/1jais.00232

Liebe, U., Glenk, K., Oehlmann, M., & Meyerhoff, J. (2015). Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys? Journal of Choice Modelling, 14, 17–31. https://doi.org/10.1016/j.jocm.2015.02.002

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. Journal of Experimental Social Psychology, 19(5), 469–479. https://doi.org/10.1016/0022-1031(83)90023-9

Mavletova, A., & Couper, M. (2015). A Meta-Analysis of Breakoff Rates in Mobile Web Surveys. En Mobile research methods: Opportunities and challenges of mobile research methodologies. Ubiquity Press. https://doi.org/10.5334/bar.f

McGuire, W. J. (1964). Some contemporary approaches to the theory of persuasion. Advances in Experimental Social Psychology, 1, 191–229.

McGuire, W. J. (1969). The nature of attitudes and attitude change. En The handbook of social psychology (2.ª ed., Vol. 3). Addison-Wesley.

McGuire, W. J. (1985). Attitudes and attitude change. En G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (3.ª ed., Vol. 2, pp. 233–346). Random House.

Montgomery, D. C. (2010). A first course in design and analysis of experiments. W. H. Freeman and Company.

Montgomery, D. C. (2017). Design and analysis of experiments. John Wiley & Sons.

Montgomery, D. C. (2019). Design and analysis of experiments (10.ª ed.). John Wiley & Sons.

Morillas Barrio, C., Munoz-Organero, M., & Sanchez Soriano, J. (2016). Can gamification improve the benefits of student response systems in learning? An experimental study. IEEE Transactions on Emerging Topics in Computing, 4(3), 429–438. https://doi.org/10.1109/TETC.2015.2497459

Mota, F., Botelho, S., & Adamatti, D. (2016). Serious games as a tool to change people attitudes: An analysis based on the discourse of collective subject. Literacy Information and Computer Education Journal, 7. https://doi.org/10.20533/licej.2040.2589.2016.0318

Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. Communication Theory, 18(3), 407–425. https://doi.org/10.1111/j.1468-2885.2008.00328.x

Muhamad, J. W., & Kim, S. (2020). Serious games as communicative tools for attitudinal and behavioral change. En H. D. O'Hair & M. J. O'Hair (Eds.), The handbook of applied communication research (1.ª ed., pp. 141–162). Wiley. https://doi.org/10.1002/9781119399926.ch9

Nan, X., Zhao, X., Yang, B., & Iles, I. (2015). Effectiveness of cigarette warning labels: Examining the impact of graphics, message framing, and temporal framing. Health Communication, 30(1), 81–89. https://doi.org/10.1080/10410236.2013.841531

Ng, B.-Y., Kankanhalli, A., & Xu, Y. (Calvin). (2009). Studying users' computer security behavior: A health belief perspective. Decision Support Systems, 46(4), 815–825. https://doi.org/10.1016/j.dss.2008.11.010

Nord, G. D., Koohang, A., Floyd, K. S., & Paliszkiewicz, J. (2020). Impact of habits on information security policy compliance. Issues in Information Systems, 21(3), 217–226. https://doi.org/10.48009/3_iis_2020_217-226

OECD. (2019). Tools and ethics for applied behavioural insights: The BASIC toolkit. OECD Publishing. https://doi.org/10.1787/9ea76a8f-en

Oehlert, G. W. (2000). A first course in design and analysis of experiments. W. H. Freeman and Company.

Parthy, P., & Rajendran, G. (2019). Identification and prevention of social engineering attacks on an enterprise. International Carnahan Conference on Security Technology (ICCST), Chennai, India, 2019, 1–5. https://doi.org/10.1109/CCST.2019.8888441

Pelling, N. (2011). The (short) prehistory of "gamification." Funding Startups and Other Impossibilities. https://nanodome.wordpress.com/2011/08/09/the-short-prehistory-of-gamification/

Perloff, R. M. (2003). The dynamics of persuasion: Communication and attitudes in the 21st century (2.ª ed.). Lawrence Erlbaum Associates.

Petty, R. E., & Cacioppo, J. T. (1986). Communication and persuasion. Springer. https://doi.org/10.1007/978-1-4612-4964-1

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. Advances in Experimental Social Psychology, 19, 123–205.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta-cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength. Social Cognition, 25(5), 657–686. https://doi.org/10.1521/soco.2007.25.5.657

Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. Journal of Personality and Social Psychology, 82(5), 722–741. https://doi.org/10.1037//0022-3514.82.5.722

Petty, R. E., Briñol Turnes, P., Fabrigar, L. R., & Wegener, D. T. (2019). Attitude structure and change. En P. A. M. Van Lange, E. T. Higgins, & A. W. Kruglanski

(Eds.), Advanced social psychology: The state of the science (pp. 117–156). Oxford University Press. https://dialnet.unirioja.es/servlet/articulo?codigo=7644599

Petty, R. E., Fazio, R. H., & Briñol, P. (2008). Attitudes: Insights from the new implicit measures (pp. xix, 544). Psychology Press. https://doi.org/10.4324/9780203809884

Petty, R. E., & Krosnick, J. A. (1995). Attitude strength: Antecedents and consequences (pp. xx, 510). Lawrence Erlbaum Associates.

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. En The handbook of social psychology (Vols. 1–2, 4.ª ed., pp. 323–390). McGraw-Hill.

Pratkanis, A. R., & Aronson, E. (1994). Age of propaganda: The everyday use and abuse of persuasion. W. H. Freeman.

Prümmer, J., Steen, T. van, & Berg, B. van den. (2025). Assessing the effect of cybersecurity training on end-users: A meta-analysis. Computers & Security, 150, 104206. https://doi.org/10.1016/j.cose.2024.104206

Rhee, H.-S., Kim, C., & Ryu, Y. U. (2009). Self-efficacy in information security: Its influence on end users' information security practice behavior. Computers & Security, 28(8), 816–826. https://doi.org/10.1016/j.cose.2009.05.008

Ripoll, O. (2014). Gamificar vol dir fer jugar. http://blogs.cccb.org/lab/article_gamificar-vol-dir-fer-jugar/

Rogers, P. (1975). A protection motivation theory of fear appeals and attitude change. The Journal of Psychology, 91(1), 93–114. https://doi.org/10.1080/00223980.1975.9915803

Rogers, P. (2014). La teoría del cambio. https://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf

Röpke, R. (2023). Extending game-based anti-phishing education using personalization: Design and implementation of a framework for personalized learning game content in anti-phishing learning games [Tesis doctoral, RWTH Aachen University]. RWTH Aachen University. https://doi.org/10.18154/RWTH-2023-04991

Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and Likert scales: Too coarse for comfort. Journal of Applied Psychology, 77(3), 336–342.

Serin, E., Handler, N., Morey, L., & Munjal, A. (2022). Randomised controlled trials: Can they inform the development of green innovation policies in the UK? LSE. https://www.lse.ac.uk/granthaminstitute/wp-content/uploads/2022/10/Randomised-control-trials_Can-they-inform-the-development-of-green-innovation-policy-in-the-UK-1.pdf

SERNAC. (2025a). Fraude en Medios de Pago en Chile. Tipologías del Fraude.

SERNAC. (2025b). Fraude en Medios de Pago en Chile. Determinantes de la Autoprotección del Consumidor y Estrategias para su Incentivo.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Sheeran, P. (2002). Intention—behavior relations: A conceptual and empirical review. European Review of Social Psychology, 12(1), 1–36. https://doi.org/10.1080/14792772143000003

Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. Social and Personality Psychology Compass, 10(9), 503–518. https://doi.org/10.1111/spc3.12265

Shrum, L. J. (2004). The cognitive processes underlying cultivation effects are a function of whether the judgments are on-line or memory-based. Communications, 29(3), 327–344. https://doi.org/10.1515/comm.2004.021

Slater, M. D., & Rouner, D. (2002). Entertainment-education and elaboration likelihood: Understanding the processing of narrative persuasion. Communication Theory, 12(2), 173–191. https://doi.org/10.1111/j.1468-2885.2002.tb00265.x

Smiderle, R., Rigo, S. J., Marques, L. B., Peçanha de Miranda Coelho, J. A., & Jaques, P. A. (2020). The impact of gamification on students' learning, engagement and

behavior based on their personality traits. Smart Learning Environments, 7(1), 3. https://doi.org/10.1186/s40561-019-0098-x

Smith, T. (2017). Gamified modules for an introductory statistics course and their impact on attitudes and learning. Simulation & Gaming, 48(6), 832–854. https://doi.org/10.1177/1046878117731888

Stajano, F., & Wilson, P. (2011). Understanding scam victims: Seven principles for systems security. Communications of the ACM, 54(3), 70–75.

Struminskaya, B., Weyandt, K., & Bosnjak, M. (2015). The effects of questionnaire completion using mobile devices on data quality: Evidence from a probability-based general population panel. Methods, Data, Analyses, 9(2), 261–292. https://doi.org/10.12758/mda.2015.014

Syafitri, W., Shukur, Z., Mokhtar, U. A., Sulaiman, R., & Ibrahim, M. A. (2022). Social engineering attacks prevention: A systematic literature review. IEEE Access, 10, 33279–33294. https://doi.org/10.1109/ACCESS.2022.3162594

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6.ª ed.). Pearson Education.

Tal-Or, N., & Cohen, J. (2010). Understanding audience involvement: Conceptualizing and manipulating identification and transportation. Poetics, 38(4), 402–418. https://doi.org/10.1016/j.poetic.2010.05.004

Tan, M., & Sagala Aguilar, K. (2012). An investigation of students' perception of Bluetooth security. Information Management & Computer Security, 20(5), 364–381. https://doi.org/10.1108/09685221211286539

Torres, I. (2021). Evidencia y explicación en economía: modelos, RCTs y su amalgama. Culturas Científicas, 2(1), 107–136. https://doi.org/10.35588/cc.v2i1.4907

Tourangeau, R., Sun, H., Yan, T., Maitland, A., Rivero, G., & Williams, D. (2018). Web surveys by smartphones and tablets: Effects on data quality. Social Science Computer Review, 36(5), 542–556. https://doi.org/10.1177/0894439317719438

Türkmen, G. P., & Soybaş, D. (2019). The effect of gamification method on students' achievements and attitudes towards mathematics. Bartın Üniversitesi Eğitim Fakültesi Dergisi, 8(1), 258–298. https://doi.org/10.14686/buefad.424575

Werbach, K. y Hunter, D. (2013). Gamificación. Revoluciona tu negocio con las técnicas de los juegos. Madrid: Pearson. N

Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. Communication Monographs, 59(4), 329–349. https://doi.org/10.1080/03637759209376276

Witte, K. (1994). Fear control and danger control: A test of the extended parallel process model (EPPM). Communication Monographs, 61(2), 113–134. https://doi.org/10.1080/03637759409376328

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J., & Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. Simulation & Gaming, 40(2), 217–266. https://doi.org/10.1177/1046878108321866

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2.ª ed.). MIT Press.

World Economic Forum. (2024). Global risks report 2024. https://www.weforum.org/publications/global-risks-report-2024/

Yasin, A., Fatima, R., Wen, L., JiangBin, Z., & Niazi, M. (2025). What goes wrong during phishing education? A probe into a game-based assessment with unfavorable results. Entertainment Computing, 52, 100815. https://doi.org/10.1016/j.entcom.2024.100815

Yasin, A., Liu, L., Li, T., Fatima, R., & Jianmin, W. (2019). Improving software security awareness using a serious game. IET Software, 13(2), 159–169. https://doi.org/10.1049/iet-sen.2018.5095

Yildirim, I. (2017). The effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. The Internet and Higher Education, 33, 86–92. https://doi.org/10.1016/j.iheduc.2017.02.002

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35(2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151

Zhao, X., Strasser, A., Cappella, J. N., Lerman, C., & Fishbein, M. (2011). A measure of perceived argument strength: Reliability and validity. Communication Methods and Measures, 5(1), 48–75. https://doi.org/10.1080/19312458.2010.547822

Zichermann, G. (2010). Fun is the future: Mastering gamification. Google Tech Talk, October 26, 2010. http://www.youtube.com/watch?v=6O1gNVeaE4g&feature=player_embedded

ANEXO 1: Tratamientos

Mensaje 1: Campaña Estándar sin refuerzo

Mensaje 2: Campaña con Mensaje Motivacional

Servicio Nacional del Consumidor

FACULTAD ECONOMÍA Y NEGOCIOS UNIVERSIDAD DE CHILE

Mensaje 3: Campaña con Explicación

Mensaje 4: Campaña con Explicación y Mensaje Motivacional

.

Anexo 2: Resultados Pruebas no paramétricas

Prueba Post Hoc Dunn: Actitud

Comparación Grupos	Zscore	P Value	P Value ajustado	r de Rank-Biserial	
1 - 2	0.530	0.596	1.000	0.008	
1 - 3	0.948	0.343	1.000	0.014	
2 - 3	0.413	0.680	1.000	0.006	
1 - 4	-0.811	0.417	1.000	0.012	
2 - 4	-1.328	0.184	1.000	0.019	
3 - 4	-1.745	0.081	1.000	0.025	
1 - 5	3.548	0.000	0.011	0.052	**
2 - 5	2.970	0.003	0.083	0.043	*
3 - 5	2.554	0.011	0.298	0.037	
4 - 5	4.342	0.000	0.000	0.063	***
1 - 6	4.364	0.000	0.000	0.064	***
2 - 6	3.763	0.000	0.005	0.055	***
3 - 6	3.341	0.001	0.023	0.049	**
4 - 6	5.167	0.000	0.000	0.075	***
5 - 6	0.752	0.452	1.000	0.011	
1 - 7	2.416	0.016	0.440	0.035	
2 - 7	1.845	0.065	1.000	0.027	
3 - 7	1.422	0.155	1.000	0.021	
4 - 7	3.226	0.001	0.035	0.047	**
5 - 7	-1.186	0.236	1.000	0.017	
6 - 7	-1.973	0.049	1.000	0.029	
1 - 8	3.784	0.000	0.004	0.055	***
2 - 8	3.199	0.001	0.039	0.047	**
3 - 8	2.781	0.005	0.152	0.041	
4 - 8	4.581	0.000	0.000	0.067	***
5 - 8	0.217	0.829	1.000	0.003	
6 - 8	-0.536	0.592	1.000	0.008	
7 - 8	1.413	0.158	1.000	0.021	

Fuente: Elaboración propia.

Prueba Post Hoc Dunn: Intención

Comparación Grupos	Zscore	P Value	P Value ajustado	r de Rank-Biserial	
1 - 2	-0.181	0.856	1.000	0.003	
1 - 3	-0.939	0.348	1.000	0.014	
2 - 3	-0.749	0.454	1.000	0.011	
1 - 4	-0.457	0.647	1.000	0.007	
2 - 4	-0.273	0.785	1.000	0.004	
3 - 4	0.477	0.633	1.000	0.007	
1 - 5	5.497	0.000	0.000	0.080	***
2 - 5	5.622	0.000	0.000	0.082	***
3 - 5	6.409	0.000	0.000	0.093	***
4 - 5	5.914	0.000	0.000	0.086	***
1 - 6	5.064	0.000	0.000	0.074	***
2 - 6	5.195	0.000	0.000	0.076	***
3 - 6	5.999	0.000	0.000	0.087	***
4 - 6	5.493	0.000	0.000	0.080	***
5 - 6	-0.566	0.572	1.000	0.008	
1 - 7	5.202	0.000	0.000	0.076	***
2 - 7	5.330	0.000	0.000	0.078	***
3 - 7	6.123	0.000	0.000	0.089	***
4 - 7	5.624	0.000	0.000	0.082	***
5 - 7	-0.349	0.727	1.000	0.005	
6 - 7	0.214	0.831	1.000	0.003	
1 - 8	4.243	0.000	0.001	0.062	***
2 - 8	4.382	0.000	0.000	0.064	***
3 - 8	5.171	0.000	0.000	0.075	***
4 - 8	4.673	0.000	0.000	0.068	***
5 - 8	-1.316	0.188	1.000	0.019	
6 - 8	-0.777	0.437	1.000	0.011	
7 - 8	-0.976	0.329	1.000	0.014	

Fuente: Elaboración propia.