

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

EVALUACIÓN DE LA CALIDAD DE SOFTWARE RESPECTO AL CRITERIO DE USABILIDAD DE LAS INTERFACES WEB DEL SISTEMA I-PAT A IMPLEMENTARSE EN EL MINISTERIO DEL INTERIOR.

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL GRADO DE
MAGISTER EN SOFTWARE CON MENCIÓN EN CALIDAD DE SOFTWARE**

DORIS KARINA TUTILLO SÁNCHEZ

doris.tutillo@epn.edu.ec

Director: PhD. Tania Calle Jiménez

tania.calle@epn.edu.ec

Codirector: PhD. Sandra Sánchez Gordón

sandra.sanchez@epn.edu.ec

Quito, junio 2019

APROBACIÓN DEL DIRECTOR

Como director y codirector del trabajo de titulación EVALUACIÓN DE LA CALIDAD DE SOFTWARE RESPECTO AL CRITERIO DE USABILIDAD DE LAS INTERFACES WEB DEL SISTEMA I-PAT A IMPLEMENTARSE EN EL MINISTERIO DEL INTERIOR, desarrollado por Doris Karina Tutillo Sánchez, estudiante de la Maestría en Software con Mención en Calidad de Software, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, damos por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa Oral.



PhD. Tania Calle
DIRECTORA



PhD. Sandra Sánchez
CODIRECTORA

DECLARACIÓN DE AUTORÍA

Yo, DORIS KARINA TUTILLO SÁNCHEZ, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normativa institucional vigente.



Ing. Doris Karina Tutillo Sánchez

DEDICATORIA

Este proyecto va dedicado a las tres personitas que llegaron a cambiar mi mundo, Dylan Sebastián, Elián Martín y mi esposo Edwin, mis tesoros, mis amores, la inspiración y motivación de mis días, quienes me han dado mucho amor y alegrías. Con quienes he formado un gran equipo para enfrenar cualquier problema y con quienes comparto los mejores momentos.

Gracias mil por estar siempre a mi lado, por regalarme sus sonrisas, sus abrazos y su amor.

Los amo con la vida.

Doris

AGRADECIMIENTO

A Dios, por las bendiciones recibidas a lo largo de este camino, por poner a las personas correctas que me han apoyado en la consecución de este objetivo.

A mi esposo Edwin, por el sacrificio, amor, apoyo y paciencia que tuvo todo este tiempo apoyándome para lograr mis metas.

A mis pequeños Dylan y Elián por luchar conmigo hombro a hombro, por comprenderme y por ser parte de este logro.

A mis papitos por su esfuerzo, su apoyo incondicional, por estar a mi lado en cada paso y por su amor con mis pequeños.

A Tania y Sandrita por su sincera amistad, por sus consejos, por compartir conmigo sus conocimientos, experiencias y guiarme en la dirección de este proyecto.

A mis hermanos Carlita, Dome, Tito y Freddy que siempre han estado pendiente de mí.

Doris

CONTENIDO

RESUMEN	1
ABSTRACT	2
1. INTRODUCCIÓN.....	3
2. METODOLOGÍA.....	12
2.1. Sistema I-Pat.....	12
2.2. Participantes.....	13
2.3. Métodos de Evaluación	14
2.3.1. Think Aloud	14
2.3.2. Focus Group	15
2.4. Cuestionarios.....	15
2.5. Ambientes y Herramientas	18
3. RESULTADOS Y DISCUSIÓN	20
3.1. Resultados Generales	20
3.1.1. Primera Iteración.....	22
3.1.2. Segunda Iteración	25
3.2. Reporte Técnico	31
3.3. Discusión.....	34
4. CONCLUSIONES.....	37
5. REFERENCIAS BIBLIOGRÁFICAS	38
6. ANEXOS.....	43
6.1. Artículo resultado de la primera iteración	43
6.2. Artículo resultado de la segunda iteración.....	43
6.3. Plantilla del cuestionario de usabilidad.....	43
6.4. Plantilla del cuestionario NAS_TLX.....	43

INDICE DE TABLAS

Tabla 1: Preguntas del Cuestionario de Usabilidad- Primera Iteración.....	16
Tabla 2: Escala de calificación de NASA-TLX [10]	18
Tabla 3: Problemas de Usabilidad con TA y FG - Primera Iteración	22
Tabla 4: Clasificación de problemas de Usabilidad según Haak	23
Tabla 5: Severidad de los problemas de usabilidad según Nielsen	23
Tabla 6: Clasificación y severidad de los problemas de usabilidad- Primera Iteración	24
Tabla 7: Clasificación y severidad de los problemas de usabilidad - Grupo Experimental.....	25
Tabla 8: Clasificación y severidad de los problemas de usabilidad - Grupo de Control.....	29
Tabla 9: Reporte Técnico con las recomendaciones para mejorar la usabilidad	33

INDICE DE FIGURAS

Figura 1: Interfaces del Sistema I-Pat.....	13
Figura 2: Proceso de evaluación de Usabilidad del aplicativo I-Pat.....	21
Figura 3: Promedio ponderado global -NASA - TLX - Usuarios del grupo experimental.....	27
Figura 4: Puntaje ponderado de la NASA-TLX - Usuarios del grupo experimental.....	28
Figura 5: Promedio ponderado global NASA - TLX - Usuarios del grupo de Control.....	30
Figura 6: Puntaje ponderado de la NASA-TLX - Usuarios del grupo de Control	31
Figura 7: Interfaz de la aplicación I-Pat antes y después de la mejora.	32

RESUMEN

La tasa de criminalidad ha aumentado en Ecuador. Las violaciones, los femicidios, los homicidios y los asaltos tienen un impacto negativo en la sociedad. Por esta razón, el gobierno ha desarrollado una aplicación móvil basada en la tecnología de Android conocida como I-Pat (Sistema Inteligente de Patrullaje) para cubrir una de las necesidades básicas de la sociedad como es la seguridad. Esta investigación resume la evaluación de la usabilidad de esta aplicación, con el uso de los métodos Think Aloud (TA) y Focus Group (FG). La combinación de estos métodos permitió cubrir aspectos de la usabilidad como consistencia, funcionalidad, navegabilidad, flujo de trabajo y control de la aplicación. Los hallazgos han dado lugar a un conjunto de recomendaciones técnicas para mejorar la funcionalidad de la aplicación y la experiencia del usuario, con el objetivo de proporcionar una aplicación que pueda ser utilizada para apoyar el trabajo de patrullaje de la Policía Nacional de Ecuador.

Palabras Clave: Think Aloud, Focus Group, Usabilidad, I-Pat, Patrullaje, calidad del producto de software, crimen

ABSTRACT

The crime rate has increased in Ecuador. Violations, femicides, homicides and assaults, are all having a negative impact on society. For this reason, the government of Ecuador has developed a mobile application based on Android technology known as I-Pat (Intelligent Police Patrolling System). This system covers one of the fundamental needs for the Ecuadorian community that is safety. This research summarizes the evaluation of the usability of this application, with the use of Think Aloud (TA) and Focus Group (FG) methods. The combination of these methods allowed covering aspects of usability such as consistency, functionality, navigability, workflow and application control. The findings have led to a number of technical recommendations being made to improve the application's functionality and the user's experience of it, and so it can be used to support the patrol work of the National Police of Ecuador.

Keywords: Think Aloud, Focus Group, I-Pat, patrolling, software quality, usability, crime

INTRODUCCIÓN

La Organización Internacional para la Estandarización (ISO / IEC 25010,2011) define la usabilidad como la capacidad del producto de software para ser comprendido, aprendido, utilizado y atractivo para el usuario, cuando se usa bajo condiciones específicas [1]. Sin embargo, este concepto no recibe la importancia adecuada, lo que conduce a confusión, error, retraso o falla absoluta para completar alguna tarea por parte del usuario. Actualmente, se están realizando múltiples esfuerzos para mejorar la seguridad, el rendimiento y la confiabilidad de los productos de software [2]. No obstante, se ha prestado poca atención a abordar los problemas de usabilidad en cualquier tipo de software.

Hoy en día, este problema está cambiando, promoviendo una evaluación de las fases iniciales del proceso de desarrollo del software y una participación más activa del usuario en las últimas fases del ciclo, incluida la usabilidad que se considera un requisito no funcional fundamental para el éxito económico en mercados competitivos [3]. Estas definiciones sugieren la posibilidad de utilizar protocolos e instrumentos específicos para medir el nivel de usabilidad en cualquier tipo de software [4].

Con el despliegue emergente de tecnologías móviles, el Ministerio del Interior decidió implementar una aplicación móvil en lugar de un sistema web. Aprovechando las ventajas que una aplicación móvil ofrece como son el acceso más rápido a la información, acceso instantáneo en línea y fuera de línea, uso de las características propias del dispositivo como son las cámaras, grabadoras de voz, escaneo de imágenes y códigos de barras. Estas ventajas apoyarán aún más el trabajo de patrullaje de los oficiales de policía. El disponer de una aplicación móvil hace necesario que se realice una evaluación de la usabilidad para garantizar que la aplicación cumpla con los estándares necesarios para que los usuarios finales cumplan sus funciones sin dificultades. Varios estudios se han centrado en la usabilidad de los dispositivos móviles [5]. Los teléfonos móviles se han convertido en un dispositivo popular en la vida cotidiana de las personas. En [6], los autores definieron cuatro dimensiones contextuales: usuario, entorno, tecnología y tarea / actividad, para considerar dentro de la evaluación de usabilidad de dispositivos móviles. Los autores en [7] aplicaron técnicas híbridas para identificar problemas de usabilidad en su trabajo de investigación. Afirmaron que la evaluación de la usabilidad debe centrarse en los problemas de uso basados en la aplicación y los errores humanos. Su estudio propuso varias dimensiones, a saber, utilidad, error,

comprensible, capacidad de aprendizaje, satisfacción e intuición. Por otro lado, las aplicaciones móviles utilizan interfaces interactivas, que aparentemente son complejas y a menudo se enfrentan a problemas de usabilidad como la sobrecarga de información, el desorden de la pantalla, la falta de soporte de tareas y los mecanismos de interacción limitados [8].

En este contexto, esta investigación realiza una evaluación de la usabilidad del Sistema Inteligente de Patrullaje de la Policía Nacional (I-Pat) desarrollado por el Ministerio del Interior en Ecuador. Este sistema cubre una de las necesidades fundamentales de la comunidad ecuatoriana, como la seguridad respaldada por tecnología, patrullaje y vigilancia de ciudades para identificar crímenes y grupos delictivos. I-Pat se basa en un modelo que genera rutas dinámicas que aplican inteligencia artificial a través del algoritmo K-means para identificar puntos críticos cuando se patrulla un circuito dentro del territorio ecuatoriano [9]. Además, I-Pat ha incorporado la herramienta API de Google Maps para diseñar rutas y verificar los medios de transporte.

En el presente estudio, la evaluación se realizó en dos iteraciones. La primera iteración se aplicó las pruebas de usabilidad a 21 estudiantes de la Facultad de Sistemas de la Escuela Politécnica Nacional utilizando Think Aloud (TA) y Focus Group (FG). La segunda iteración la evaluación se realizó con profesionales en Ciencias de la Computación y con miembros activos de la Policía Nacional de Ecuador utilizando TA y FG para la evaluación de usabilidad y NASA-TLX para la medición de la carga de trabajo.

Actualmente, esta aplicación se encuentra en la fase de prueba, debido a las evaluaciones necesarias para apoyar las actividades diarias durante las tareas de patrullaje y también al nivel de facilidad de uso necesario para que los miembros de la Policía utilicen la aplicación sin dificultades. De lo contrario, la aplicación no se puede implantar.

La usabilidad de la aplicación móvil se evaluó con los métodos TA y FG que se utilizaron para proponer una evaluación que enfatiza aspectos como la estructura global del sitio, el enfoque en el contenido y la navegabilidad, sin una terminología compleja y con el objetivo de guiar al usuario por un camino fácil de entender, Además, se agregó NASA-TLX como dimensión adicional de usabilidad a la aplicación móvil. Esta prueba evalúa la experiencia subjetiva del usuario de la carga de trabajo general y los factores que contribuyen a ella [10].

TA como un método de evaluación de usabilidad se utiliza para obtener información sobre cómo las personas trabajan con un producto o una interfaz [11]. Se trata de hacer que un usuario verbalice sus procesos de pensamiento mientras realiza una tarea o un conjunto de tareas. A menudo, un observador está presente para inducir al usuario a ser más vocal en su trabajo. Es útil para identificar problemas y relativamente fácil de configurar. Además, puede dar una idea de la actitud del usuario, que normalmente no se puede discernir a partir de una encuesta o cuestionario [12].

FG es un método común utilizado en la investigación de mercados para determinar las actitudes de los clientes sobre los productos o servicios [13]. Para los participantes, la sesión de FG debe sentirse libre y relativamente desestructurada, pero en realidad, el moderador debe seguir un guion pre planificado de problemas específicos y establecer metas para el tipo de información que se recopilará [14]. Este método consiste en una discusión planificada y diseñada sobre las interfaces. En esta investigación, la discusión fue guiada y facilitada por un moderador. El moderador siguió una estructura predefinida para que la discusión permaneciera enfocada. La actividad que se utilizó fue una lluvia de ideas para que los participantes intercambiaran opiniones.

NASA-TLX es una técnica de evaluación de la carga de trabajo subjetiva que se basa en una construcción multidimensional para derivar una carga de trabajo general. La puntuación se basa en un promedio ponderado de calificaciones en seis sub escalas: demanda mental, demanda física, demanda temporal, rendimiento, esfuerzo y nivel de frustración. Tres de las sub escalas se relacionan con las demandas impuestas al usuario (demanda mental, física y temporal), mientras que las otras sub escalas se enfocan en la interacción del uso con la tarea (rendimiento, esfuerzo y nivel de frustración) [15]. La NASA TLX se ha utilizado en varios campos [16], como estudios que incluyen la evaluación de pantallas visuales y auditivas, dispositivos de entrada vocal y manual, y visión virtual y aumentada. Las investigaciones han explorado la relación entre las calificaciones TLX de la NASA y otros factores de rendimiento, como fatiga, estrés, confianza, experiencia y conocimiento de la situación. Debido a su amplia aplicabilidad, también se ha utilizado en combinación con otros métodos para la evaluación de la usabilidad [17].

La combinación de TA y FG como pruebas de usabilidad es un enfoque factible para identificar aspectos importantes de usabilidad y para abrir la participación del usuario final con la aplicación. El método TA fue útil para identificar los requisitos detallados relacionados con la

interfaz de usuario. Las sesiones de FG, por otro lado, fueron importantes para incorporar temas contextuales. Además, la aplicación de NASA – TLX permitió medir la estimación de la carga de trabajo sensible a la complejidad de las tareas que se evaluaron a los usuarios finales. Finalmente, se desarrolló un informe que presenta los errores de usabilidad encontrados con sus respectivas recomendaciones técnicas. Este informe fue entregado a la Policía Nacional de Ecuador para la posterior mejora del Sistema I-Pat.

1. REFERENCIAL TEÓRICO

Para el desarrollo de este proyecto es importante realizar una revisión de literatura relacionada con los métodos de evaluación de la usabilidad.

1.1. REVISIÓN DE LA LITERATURA

Para el presente trabajo de titulación, se realizó una revisión de literatura basada en la propuesta metodológica descrita en [18] por Kitchenham, la cual consiste en tres fases:

fase de búsqueda, fase de ejecución y fase de extracción

1.1.1. Fase de Búsqueda

En la fase de búsqueda se definen las preguntas de investigación, estas preguntas son la guía bajo el cual se realiza el estudio y clasificación del material bibliográfico. Las preguntas de investigación definidas para este trabajo están descritas en la Tabla 1

Tabla 1: Preguntas de Investigación

RQ	Preguntas de investigación
1	¿Qué métodos de evaluación de la usabilidad existen para aplicaciones móviles?
2	¿Cuáles son las ventajas de una evaluación de usabilidad con TA y FG?
3	¿Qué recomendaciones existen para la combinación de métodos y técnicas?

El siguiente punto que se realiza en esta fase es la definición de cadenas de búsqueda

Las cadenas de búsqueda se diseñaron en función y combinación de los términos derivados de las preguntas de investigación dando como resultado:

- Usability test AND mobile
- Usability evaluation methods
- Think Aloud Usability method
- Focus Group Usability method

Luego de las cadenas de búsqueda, se definen las fuentes de información para el presente trabajo las fuentes de información consideradas fueron Google Scholar, The ACM Digital Library, IEEE Xplore, ResearchGate, Elsevier y Springer.

1.1.2. Fase de ejecución

Se ejecutan las búsquedas en base a las cadenas definidas en la fase anterior. Al momento de extraer los resultados, estos no siempre son un aporte para la investigación que se está realizando, por lo cual es necesario definir criterios de inclusión y exclusión para poder delimitar el umbral de la investigación.

Los criterios de inclusión definidos para este trabajo son:

- Período de tiempo: Se define un intervalo temporal de publicación entre los años 2014 y 2019, con la finalidad de identificar la mirada actual del problema planteado.
- Idioma: Se toman en cuenta publicaciones en inglés, debido a que existen más trabajos realizados en este idioma.
- Contenido: Se consideran artículos que describan tres puntos clave para esta investigación, los cuáles son: Evaluación de la usabilidad, Think Aloud, Focus Group.
- Los criterios de exclusión definidos para este trabajo son:
 - Contenido: Excluir publicaciones no relacionadas con el dominio de investigación

1.1.3. Fase de extracción

En esta fase, se analizan cada uno de los artículos encontrados para poder identificar los que serán un aporte para la investigación. El análisis comienza por el título del artículo, verificando que este enmarcado dentro de la búsqueda que se está realizando. El siguiente punto a analizar es el abstract del artículo, al leer el abstract del artículo podemos validar que sea un aporte para el estudio que se está realizando.

Inicialmente se encontraron 63 trabajos relacionados a las preguntas de investigación, luego de aplicar los criterios de exclusión y elegir los más relevantes quedaron en 34.

La Tabla 2 muestra el número de artículos encontrados por cada base de datos científica.

Tabla 2: Resultados Revisión de la Literatura

Base de datos Científicas	Total de artículos
ACM Digital Library	8
Google Scholar	13
IEEE Xplore	4
ResearchGate	6
Springerlink	2
ScienceDirect	1
Total	34

1.1.4. Resultados de la revisión de la Literatura

En la revisión de la literatura, los autores han identificado el uso de TA y FG como métodos para medir la usabilidad de varios productos de software. [11] define TA como un método de evaluación de usabilidad utilizado para obtener información sobre cómo las personas trabajan con un producto o interfaz. [19] sostiene que los profesionales de la usabilidad actualmente usan variaciones de la AT como el método principal para identificar problemas de usabilidad. Este estudio compara tres Protocolos TA para usar en pruebas. El primer protocolo, llamado Instrucción, es donde no hay instrucciones para cantidades variables de instrucción o sesiones de práctica, luego Intervención: diferentes tipos de intervención con una variedad de sondeos o indicaciones, y Solicitud: tasas variables de la prueba con administradores que administran un rápido. Este estudio muestra las variaciones de TA con estos protocolos que los profesionales de la usabilidad utilizan actualmente, los profesionales que desean medidas que reflejen el comportamiento del usuario sin ayuda pueden elegir entre dos métodos: tradicional o comunicación por voz, que no difieren significativamente entre sí en términos de sus efectos en Rendimiento del usuario. En el estudio [20] se utilizó el método de pensamiento en voz alta concurrente. Esta es una técnica muy valiosa que se utiliza para comprender los procesos cognitivos basados en tareas de las personas,

tanto en las pruebas de usabilidad como en el estudio más amplio de la interacción persona-computadora.

Además, [21] mostró que un pensamiento en voz alta relajado, que incluyó intervenciones del evaluador, aumentó el tiempo de prueba y condujo a cambios en el comportamiento en la interfaz. [22] tiene otros enfoques para recopilar datos verbales; estos incluyen la interacción constructiva, donde los usuarios trabajan juntos en parejas. La interacción constructiva puede tomar varias formas: los usuarios pueden trabajar juntos en pie de igualdad [23] o se pueden usar formatos alternativos, tales como escenarios de aprendizaje, en los que el usuario enseña a otra persona sobre el sistema de enfoque. En [24] comentarios, los usuarios dijeron que podrían no dar razones para sus acciones durante el pensamiento en voz alta, solo una descripción de las acciones en sí.

[25] propone algunas herramientas para capturar las observaciones durante las sesiones usando TA. Otros estudios se han centrado más en las preguntas tácticas de la evaluación de la usabilidad, por ejemplo, [26] presenta un ejemplo de un informe de resultados que se debe presentar después de la evaluación con TA.

La literatura sugiere que la TA es un método ampliamente utilizado para evaluar la usabilidad tomando su versión inicial propuesta por Ericson y Simon o con diferentes versiones que obtienen buenos resultados, como lo muestra [27], donde se descubrió que los pensamientos en voz alta se clasificaron entre los cinco Principales métodos utilizados por los profesionales de la usabilidad en Suecia en prototipos, entrevistas, estudios de campo y escenarios.

Por otro lado, los FG se han utilizado durante décadas como se menciona en [28]. La discusión se usa frecuentemente como un enfoque cualitativo para obtener una comprensión profunda de los problemas sociales. El método apunta a obtener datos de un grupo de individuos seleccionados deliberadamente en lugar de una muestra estadísticamente representativa de la población más amplia. Los FG son menos amenazadores para muchos participantes en la investigación, y este ajuste es útil para que los participantes discutan las percepciones, ideas, opiniones y pensamientos. Por ejemplo, [29] propone una serie de pasos para maximizar la confiabilidad de los datos y recomienda algunos criterios como confiabilidad, validez interna, transferibilidad y

confiabilidad para asegurar que se informen aspectos importantes de la investigación cualitativa, lo que aumenta la calidad de los estudios publicados de FG.

[30] especifica un marco para capturar y analizar la información derivada de los grupos de interés en relación con el tipo de información generada, los métodos cuasi estadísticos y el análisis de conversaciones. Además [31] propone que las FG puedan proporcionar una experiencia empírica valiosa y complementaria de manera rápida y a bajo costo. [32] utiliza el método para identificar las características de usabilidad con respecto a un sitio web institucional.

Sin embargo, como se menciona en [14], aunque los grupos focales pueden ser una herramienta poderosa en el desarrollo de software, no se debe usar como la única fuente de datos de usabilidad. [33] revisa algunas de las características de las técnicas de usabilidad y considera la AT como un complemento en la aplicación de los GF. En consecuencia, la presente investigación utilizará TA y FG como métodos de evaluación de usabilidad para el sistema I-Pat.

2. METODOLOGÍA

Al considerar la seguridad como una necesidad de la sociedad global, este estudio aplica los métodos de evaluación de TA y FG, además de la aplicación de cuestionarios de usabilidad y cuestionario NASA-TLX al sistema de patrullaje de la Policía desarrollado por el Ministerio del Interior del Ecuador. Uno de los objetivos de la aplicación de estos métodos es obtener errores de usabilidad según las características de sus participantes, como su comportamiento, género, edad y conocimiento de los sistemas móviles.

El sistema I-Pat, los participantes en la evaluación y el proceso de evaluación se describen a detalle a continuación.

2.1. Sistema I-Pat

I-Pat es el acrónimo de Intelligent Police Patrolling. Este sistema cubre una de las necesidades fundamentales de la comunidad ecuatoriana, es decir, la seguridad. Con el uso de la tecnología, el patrullaje y la vigilancia de las ciudades se facilitarán mediante la identificación de delitos y grupos delictivos, ayudando al trabajo de la Policía Nacional. En este contexto, I-Pat fue desarrollado para aplicaciones que utilizan el sistema Android para realizar patrullajes inteligentes para la Policía Nacional del Ecuador. Para este propósito, utiliza información sobre los delitos denunciados en tres circuitos durante un período de cuatro años, respaldado por un modelo que genera rutas dinámicas que aplican inteligencia artificial con el algoritmo K-means para identificar puntos críticos al patrullar un Circuito dentro del territorio ecuatoriano. Además, la herramienta API de Google Maps se utiliza para diseñar las rutas y los medios de transporte que utilizará el oficial de policía. El sistema I-Pat tiene módulos históricos que detallan la información actual disponible sobre robo de personas, robo de casas, robo de bancos, robo de vehículos, robo de motocicletas, robo de accesorios y robo en carreteras. También tiene una sección complementaria con funciones que incluyen solicitudes de patrullas policiales de refuerzo, mapas predictivos, horarios de los agentes y disturbios civiles. La Figura. 1 muestra ejemplos de las interfaces de los módulos de patrulla, disturbios civiles y robo de vehículos.

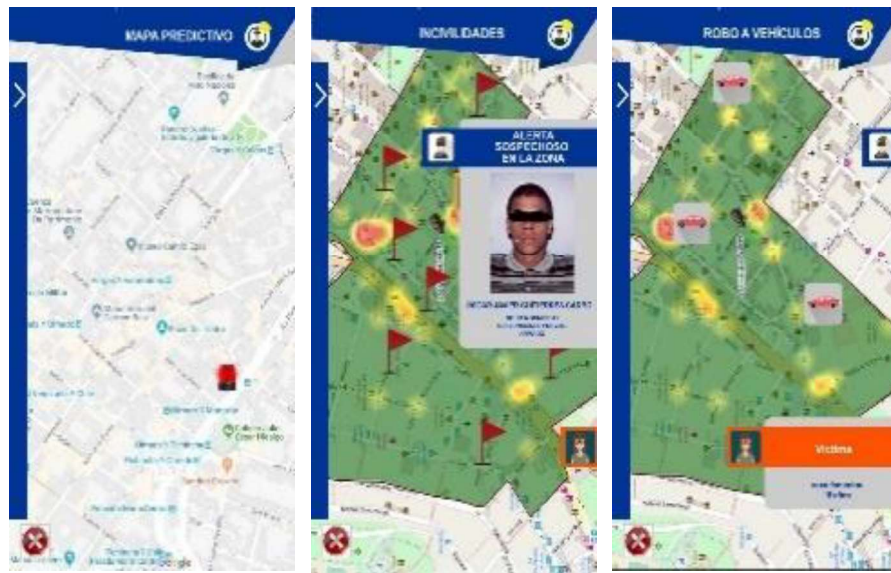


Figura 1: Interfaces del Sistema I-Pat

2.2. Participantes

En esta investigación, se evaluó la usabilidad del Sistema I-Pat trabajando en dos iteraciones.

Primera Iteración:

Se contó con la participación de estudiantes de la Escuela Politécnica Nacional de Ecuador. Los participantes en este estudio consistieron en dos grupos, incluidos evaluadores y facilitadores. Los evaluadores incluyeron a veintidós estudiantes de ingeniería informática en su sexto y séptimo semestre que están capacitados para realizar la evaluación de la calidad de diferentes sistemas, ya que los estudiantes han aprobado el curso de calidad del software. Los evaluadores tenían conocimientos sobre el diseño y el análisis de sistemas informáticos, así como sobre los principios y conceptos de los lenguajes de programación, pero no tenían conocimientos ni experiencia en el uso de I-Pat

Segunda Iteración:

En esta iteración se trabajó con dos grupos, el primero denominado grupo de experimentación que estaba constituido por profesionales en el área de Sistemas con

conocimientos sobre el diseño y el análisis de sistemas informáticos, así como sobre los principios y conceptos de los lenguajes de programación y de la calidad de software.

El segundo grupo denominado grupo de control estuvo conformado acorde al principio de Nielsen por 5 miembros de la Policía Nacional del Ecuador. Nielsen [34] afirma que los mejores resultados provienen de las pruebas de no más de 5 usuarios y la ejecución de tantas pruebas pequeñas como pueda permitirse, porque después del quinto usuario, se pierde tiempo al observar los mismos hallazgos repetidamente pero no se aprende sobre nuevos errores. Este grupo de usuarios tenían conocimiento básico del manejo de sistemas de información y amplio conocimiento en tareas de patrullaje, información delictiva, manejo de criminales y en sí del manejo de información policial.

Antes de comenzar la evaluación, en ambas iteraciones todos los usuarios estaban familiarizados verbalmente con los objetivos del estudio y se obtuvo su consentimiento para realizar las pruebas. Para la aplicación del primer método de con TA, el facilitador recordó a los usuarios que expresaran sus pensamientos en voz alta. Mientras que, para el segundo método FG, el facilitador asumió el papel de moderador, y ordenó a los grupos obtener el mayor número de ideas con respecto a la usabilidad de I-Pat.

2.3. Métodos de Evaluación

2.3.1. Think Aloud

El método TA tiene sus raíces en la investigación psicológica. Fue desarrollado a partir del antiguo método de introspección. La introspección se basa en la idea de que uno puede observar eventos que tienen lugar en la conciencia, como uno puede observar eventos en el mundo exterior [35]. El método TA como herramienta para respaldar las pruebas de usabilidad se ha clasificado entre los principales métodos utilizados por los profesionales de la usabilidad sobre prototipos, entrevistas, estudios de campo y escenarios [27]

El uso frecuente de TA es porque no es posible observar directamente lo que piensa un usuario. Con el uso de TA, las áreas donde un usuario está teniendo dificultades y las razones de las dificultades se expresan verbalmente. El evaluador de usabilidad utiliza esta información junto con otras métricas para identificar las áreas problemáticas

de la aplicación que se está evaluando y para diseñar sugerencias de mejora. Además, se alienta a los participantes a pensar en voz alta mientras trabajan en una tarea.

2.3.2. Focus Group

FG surgió como un método de investigación en la década de 1950 en la investigación social cuando los investigadores ampliaron el formato de entrevista abierta a una discusión grupal [36]. FG es una forma de recopilar datos cualitativos, que esencialmente involucra la participación de un pequeño número de personas en una discusión grupal informal, centrada en un tema específico o conjunto de problemas [37].

FG se usa frecuentemente como un enfoque cualitativo para obtener una comprensión profunda de los problemas sociales. El método apunta a obtener datos de un grupo de individuos seleccionados deliberadamente en lugar de una muestra estadísticamente representativa de una población más amplia. FG incluye discusiones cuidadosamente planificadas, diseñadas para obtener las percepciones de los miembros del grupo en un área de interés definida. Normalmente hay entre tres y 12 participantes y la discusión es guiada y facilitada por un moderador, que sigue una estructura predefinida para que la discusión permanezca enfocada. Los miembros se seleccionan en función de sus características individuales según el tema de la sesión. El entorno grupal permite a los participantes desarrollar las respuestas e ideas de los demás, lo que aumenta la riqueza de la información obtenida [31].

2.4. Cuestionarios

Primera Iteración

Después de la evaluación, los usuarios llenaron un cuestionario que consta de 20 preguntas para medir la facilidad de uso. Este cuestionario se basó en QUIS [38], SUMI [39] y SUS [40].

El cuestionario de la Escala de usabilidad del sistema SUS es uno de los cuestionarios más conocidos para combinar el número reducido de preguntas y la precisión [41], mide la facilidad de uso de una herramienta, un programa informático, un instrumento, etc. Está compuesto por diez elementos evaluados en una escala Likert de uno a cinco, donde uno está totalmente en desacuerdo y estoy totalmente de acuerdo, necesita poco tiempo para

responder, es fácil de calificar y es fácilmente comparable con otros instrumentos. Se recomienda aprobarlo una vez que los usuarios hayan trabajado con la aplicación o herramienta que se va a evaluar.

El cuestionario SUMI -Inventario de Medición de Usabilidad de Software. Es un cuestionario de pago que mide la usabilidad de un software o servicio, que se puede utilizar para evaluar nuevos productos, hacer comparaciones de versiones anteriores para alcanzar nuevos y mejores desarrollos, consta de 50 elementos para evaluar en una escala de uno a tres (desacuerdo, no sé y estoy de acuerdo).

El cuestionario QUIS -Cuestionario para la satisfacción de la interfaz de usuario- permite evaluar la satisfacción de los usuarios. Los ítems se presentan en una escala de cero a nueve, siendo cero confuso y nueve claros. Consiste en cinco categorías de preguntas: 1) Reacciones generales sobre el software, 2) ventanas, 3) terminología e información del sistema, 4) aprendizaje y 5) capacidades del sistema.

El cuestionario aplicado en esta iteración consistió de cuatro secciones, la primera contenía preguntas sobre detalles demográficos, como la edad, el género y la educación. También se centró en la experiencia de los participantes en el trabajo con dispositivos móviles. La segunda sección se enfocó en aprender sobre el sistema, y la tercera sección trabajó en temas relacionados con la capacidad del sistema, la sección final trabajó con usabilidad e interfaces de usuario. Los ítems se presentan en una escala de cero a nueve, siendo cero confuso y nueve claros. La tabla 1 muestra los ítems del cuestionario.

Tabla 3: Preguntas del Cuestionario de Usabilidad- Primera Iteración

<i>Item</i>	<i>Descripción</i>
Primera sección: Datos Demográficos	
1	Edad
2	Género
3	Experiencia con sistemas de computación
4	Experiencia con el manejo de dispositivos móviles.
Segunda sección: Aprendizaje	
5	Manejo del sistema
6	Exploración de nuevas características por prueba y error
7	Recuerda nombres y uso de comandos
8	Las tarea se realizan de manera rápida y sencilla

9	La herramienta fue fácil de usar.
10	Piensa que la mayoría de la gente aprenderá a usar la herramienta muy rápidamente
11	En un inicio aprender a utilizar esta herramienta es difícil
<hr/>	
Tercera Sección: capacidades del Sistema	
12	Velocidad del sistema
13	Confiabilidad del sistema
14	El software en algún momento se detuvo inesperadamente
<hr/>	
Cuarta Sección: usabilidad e interfaz de usuario	
15	Uso de colores
16	Retroalimentación del sistema
17	Trabajar con este software es mentalmente estimulante
18	La forma en que el sistema presenta la información es clara y comprensible
19	Puedo comprender y manejar la información proporcionada por el sistema
<hr/>	
Comentarios	

Segunda Iteración

En esta iteración se aplicó el cuestionario NASA-TLX. Nasa-TLX es un procedimiento de evaluación multidimensional, en el que se obtiene un puntaje de carga de trabajo global basado en el promedio ponderado de seis variables o sub escalas que son: demandas mentales, demandas físicas, demandas temporales, esfuerzo, frustración y rendimiento. El concepto de la NASA-TLX es permitir que el usuario acceda a la situación de la carga de trabajo una vez que finalice la prueba. Las sub escalas proporcionan información detallada, no solo preguntas de una respuesta. NASA-TLX no prueba al usuario mientras está en el proceso de completar la tarea. El usuario se ve obligado a confiar en lo que recuerda y proporciona, y a la opinión basada en la memoria [42].

La evaluación tiene dos fases. En la primera fase, los usuarios asignan peso a cada dimensión antes de la ejecución de la tarea. En la segunda fase, los usuarios asignan una puntuación a cada una de las seis dimensiones inmediatamente después de la ejecución de la tarea [43]. El objetivo principal era predecir la carga de trabajo generada por las interfaces web, agregando así una dimensión más de usabilidad a la aplicación móvil y garantizando la mayor seguridad del usuario.

La NASA-TLX proporciona una puntuación general basada en un promedio ponderado de calificaciones de las seis sub escalas: demanda mental, física y temporal, esfuerzo, rendimiento y nivel de frustración. La tabla 1 describe cada uno de ellos.

Tabla 4: Escala de calificación de NASA-TLX [10]

Dimensión	Extremos	Descripción
Exigencia mental (m)	Baja/alta	¿Cuánta actividad mental y perceptiva fue necesaria? (Por ejemplo: pensar, decidir, calcular, recordar, buscar, investigar, etc.). ¿Se trata de una tarea fácil o difícil, simple o compleja, pesada o ligera ?
Exigencia física (f)	Baja/alta	¿Cuánta actividad física fue necesaria? (Por ejemplo: empujar, tirar, girar, pulsar, accionar, etc.) ¿Se trata de una tarea fácil o difícil, lenta o rápida, relajada o cansada?
Exigencia temporal (t)	Baja/alta	¿Cuánta presión de tiempo sintió, debido al ritmo al cual se sucedían las tareas o los elementos de las tareas? ¿Era el ritmo lento y pausado o rápido y frenético?
Esfuerzo (e)	Bajo/alto	¿En qué medida ha tenido que trabajar (física o mental mente) para alcanzar su nivel de resultados?
Rendimiento (r)	Bueno/malo	¿Hasta qué punto cree que ha tenido éxito en los objetivos establecidos por el investigador (o por Ud. mismo)? ¿Cuál es su grado de satisfacción con su nivel de ejecución?
Nivel de frustración (fr)	Bajo/alto	Durante la tarea, en qué medida se ha sentido inseguro, desalentado, irritado, tenso) o preocupado o por el contrario, se ha sentido seguro, contento, relajado y satisfecho ?

2.5. Ambientes y Herramientas

La evaluación con en la primera iteración y con el grupo experimental en la segunda iteración se realizaron en la Escuela Politécnica Nacional de Ecuador. Para el grupo de control de la segunda iteración, la evaluación se llevó a cabo en las instalaciones del Regimiento Mariscal Sucre de la Policía Nacional del Ecuador. El proceso de evaluación se llevó a cabo con el software de grabación DU utilizado para capturar la interacción del usuario con la aplicación móvil I-Pat. El micrófono y la cámara de video se utilizaron para grabar la voz del usuario y capturar los gestos y modos del usuario, respectivamente. El

proceso de evaluación incluyó dos tareas: 1) Navegación histórica, donde se muestra la información de los actos criminales llevados a cabo en la ciudad, 2) El módulo de complementos que muestra la información relacionada con las patrullas, los horarios del personal policial y la concentración del delito.

3. RESULTADOS Y DISCUSIÓN

3.1. Resultados Generales

La Figura 2 muestra el proceso seguido para aplicar TA y FG para evaluar la usabilidad de la aplicación móvil I-Pat y NASA_TLX para medir la carga de trabajo. Una vez que los evaluadores iniciaron la aplicación, se realizó una introducción a las funciones del sistema para familiarizar a los usuarios.

La primera sesión utilizó TA, el método que permite recopilar información sobre la interacción cognitiva de los usuarios con el sistema [44]. Se pidió a los usuarios que expresaran lo que vieron, pensaron y sintieron con respecto al uso de las interfaces I-Pat. Después de esta introducción, el TA duró 15 minutos. Durante esta sesión, el facilitador les recordó a los usuarios que expresaran sus pensamientos en voz alta y los grabaran en audio para su posterior análisis.

La segunda sesión se trabajó con el método FG, para lo cual fue necesario seguir las recomendaciones en [45], donde se indica la importancia de determinar el tamaño del grupo focal, la cantidad de preguntas formuladas, el tiempo asignado para cada pregunta, el formato de la sesión del grupo focal y la duración de la sesión para tener éxito en la aplicación de este método. El moderador realizó una lluvia de ideas con los participantes con el objetivo de obtener ideas originales sobre la evaluación de la usabilidad de la aplicación móvil, a través del intercambio libre de ideas entre cada uno de los evaluadores. El moderador explicó las reglas de la sesión, como evitar las críticas y expresar todas las ideas que los usuarios tienen. Además, el moderador mencionó que todos los pensamientos eran importantes, incluso si esos pensamientos no parecían tener sentido común, ya que esas ideas podían servir de inspiración para el desarrollo de nuevas propuestas. Los probadores expresaron sus ideas al resto del grupo.

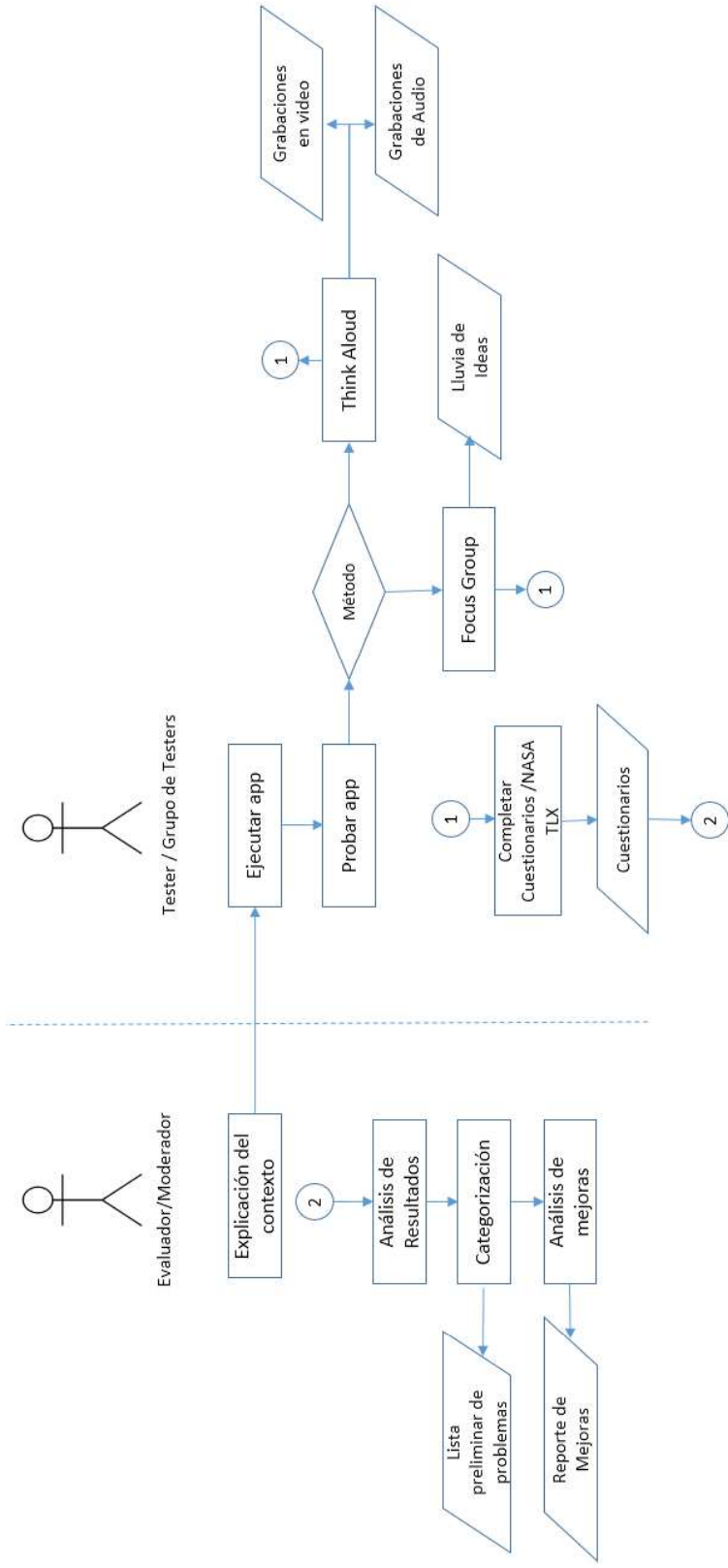


Figura 2: Proceso de evaluación de Usabilidad del aplicativo I-Pat

3.1.1. Primera Iteración

En este primer estudio [46], los participantes incluyeron trece (62%) hombres y ocho (38%) mujeres con rangos de edad entre 18-22 (66.7%), 23-27 (28.6%) y 28-32 (4.8%). Se encontraron dieciocho errores con el método TA; sin embargo, los errores fueron reportados varias veces por diferentes usuarios por lo que agrupando a los errores bajo un mismo nombre se obtuvieron nueve.

Por otra parte, se informaron 34 errores con el método FG, con este método se mantienen los errores de funcionalidad y la falta de ayuda que se presentó al aplicar TA y se agregan errores de interfaz, de igual forma se agrupan los errores bajo un mismo nombre y se obtienen 13 errores. La tabla 3 muestra los resultados obtenidos tras la aplicación de TA y FG respectivamente.

Tabla 5: Problemas de Usabilidad con TA y FG - Primera Iteración

Ítem	Problemas	TA	FG
1	Secciones sin ayuda	x	
2	Términos no intuitivos	x	x
3	Botones sin etiquetas	x	x
4	El zoom en el mapa no funciona	x	x
5	Botón "cerrar" mal ubicado	x	x
6	Botones sin funcionalidades	x	x
7	Patrullaje no función	x	
8	Distribución incorrecta del menú	x	
9	Mapas sin ayuda	x	
10	No se permite personalización		x
11	No se muestran mensajes		x
12	No existe manual de usuario		x
13	Botones sin ayuda	x	x
14	Colores no adecuados		x
15	El menú no trabaja bien		x

Para resumir, se encontraron cincuenta y dos errores reportados por veintiún usuarios, pero, después del análisis y la categorización, se resumieron a quince errores. Los errores encontrados se clasificaron según el método de Haak [47], este método establece cuatro categorías principales descritas en la Tabla 4.

Tabla 6: Clasificación de problemas de Usabilidad según Haak

Ítem	Categoría	Descripción
1	Problemas de interfaz	El participante no puede detectar un elemento en particular en una pantalla.
2	Problemas de terminología	El participante no comprende parte (s) de la terminología.
3	Problema de entrada de datos	El participante no sabe interactuar con la búsqueda..
4	Problemas de comprensión	El sistema carece de la información necesaria para utilizarlo eficazmente..

Además, se utilizó el método de Nielsen [48] para identificar la severidad de los problemas encontrados según la clasificación en la Tabla 5.

Tabla 7: Severidad de los problemas de usabilidad según Nielsen

Ítem	Problemas
0	No existe problema de usabilidad
1	Problema cosmético
2	Problema de usabilidad menor
3	Problema de usabilidad mayor
4	Catástrofe de usabilidad

La Tabla 6 muestra la clasificación y la severidad de los problemas que los usuarios encontraron en el proceso de evaluación. De acuerdo con la tabla, los problemas de Comprensión fueron los más frecuentes con siete (46%) de los casos. Alrededor del 43% de los problemas en esta categoría tenían una severidad de 2 y 3. En esta categoría, los principales problemas reportados fueron el mal funcionamiento del zoom en los mapas, funciones como la patrulla que no funciona, iconos no descriptivos, funciones cuyo propósito era desconocido. Luego, los usuarios identificaron seis (40%) problemas de diseño. En esta categoría, cuatro problemas se clasificaron como severidad 2, los problemas relacionados con la falta de ayuda en los iconos, la ausencia de mensajes, los colores inapropiados. Estos fueron los que más se repitieron.

Tabla 8: Clasificación y severidad de los problemas de usabilidad- Primera Iteración

Ítem	Problemas	Categoría	Severidad	Método
1	Zoom en el mapa no funciona	Problema de entrada de datos	3	TA/FG
2	Distribución incorrecta del menú	Problemas de interfaz	1	TA
3	Iconos sin descripción	Problemas de comprensión	2	FG
4	Botones sin funcionalidad	Problemas de comprensión	3	TA/FG
5	Títulos de los botones sin sentido	Problemas de comprensión	3	TA
6	Mapa sin ayuda	Problemas de comprensión	3	TA
7	Botones mal ubicados		1	TA/FG
8	Colores inadecuados	Problemas de interfaz	1	FG
9	El menú no trabaja bien	Problema de entrada de datos	3	FG
10	Personalización no permitida	Problemas de interfaz	1	FG
11	Mensajes no se muestran	Problema de entrada de datos	3	FG
12	Funciones no intuitivas	Problemas de comprensión	3	FG
13	Falta de ayuda para el usuario	Problemas de comprensión	2	TA/FG
14	Patrullaje no funciona	Problemas de comprensión	3	TA
15	Términos no intuitivos	Problemas de terminología	2	TA/FG

Los resultados principales, después de la aplicación del cuestionario, fueron con respecto a la experiencia de los usuarios con el funcionamiento del sistema. De

acuerdo con esto, quince (71%) usuarios no tuvieron problemas al usar la aplicación y cuatro (19%) casos tuvieron una o tres interrupciones sin detener el proceso de evaluación.

Además, dos (9%) problemas quedaron sin resolver después de la revisión del usuario, y el proceso de evaluación continuó sin esa tarea. Después de las evaluaciones, el 90% de los usuarios hicieron propuestas para mejorar el sistema. La mayoría de las sugerencias se relacionaron con un mejor diseño (57%) y mejoras con las funciones que usa Google Maps (21%).

3.1.2. Segunda Iteración

En el grupo experimental, los participantes incluyeron sesenta (60%) mujeres y cuarenta (40%) hombres con rangos de edad entre 19-27 (25%), 28-34 (35%), más de 32 (40%). Veintiún errores fueron encontrados con el método TA. Sin embargo, los errores fueron reportados varias veces por diferentes usuarios. La mayor parte de los errores encontrados en TA estaban relacionados con términos incorrectos o no intuitivos, la falta de ayudas textuales o un manual del usuario. Además, se reportaron errores de funcionalidad. Por otro lado, se informaron 34 errores con el método FG, con este método se mantienen los errores de funcionalidad y la falta de ayuda encontrada con el TA y se agregan errores de interfaz. Para resumir, se encontraron cincuenta y dos errores informados por dieciocho usuarios, pero, después del análisis y la categorización, se resumieron en dieciocho errores [49].

Tabla 9: Clasificación y severidad de los problemas de usabilidad - Grupo Experimental

N°	Problemas	Categoría	Severidad	Método
1	Zoom en los mapas no está funcionando.	Problema de entrada de datos	3	TA/FG
2	Distribución incorrecta del menú	Problemas de interfaz	1	TA
3	Iconos sin descripción	Problemas de comprensión	2	FG
4	Botones sin funcionalidades	Problemas de comprensión	3	TA/FG

5	Títulos en los botones sin sentido	Problemas de comprensión	3	TA
6	Mapas sin ayuda	Problemas de comprensión	3	TA
7	Botones mal ubicados	Problemas de interfaz	1	TA/FG
8	Colores no adecuados	Problemas de interfaz	1	FG
9	Opción de menú no funciona adecuadamente	Problema de entrada de datos	3	FG
10	No se permite personalización	Problemas de interfaz	1	FG
11	No se muestran mensajes	Problema de entrada de datos	3	FG
12	Funciones poco intuitivas	Problemas de comprensión	3	FG
13	No hay ayuda para el usuario	Problemas de comprensión	2	TA/FG
14	Opción de patrullaje no funciona	Problemas de comprensión	3	FG
15	Términos poco intuitivos	Problemas de terminología	2	TA/FG
16	Superposición de menús	Problema de entrada de datos	2	TA/FG
17	Mapas sin leyendas	Problemas de terminología	1	TA
18	Restricciones de Sistema operativo	Problema de entrada de datos	1	TA/FG

La tabla 7 muestra la clasificación y la severidad de los errores que los usuarios encontraron en el proceso de evaluación para los usuarios del grupo experimental. Según la Tabla 2, los problemas de comprensión fueron los más frecuentes con siete (39%) de los casos. Alrededor del 71% de los problemas en esta categoría tenían una severidad de tres. En esta categoría, los principales problemas reportados fueron funciones como la patrulla que no funciona, íconos no descriptivos, funciones cuyo propósito era desconocido. Luego, los usuarios identificaron cinco (28%) problemas de entrada de datos. En esta categoría, tres problemas se clasificaron como severidad 3, el mal funcionamiento del zoom en los mapas, los mensajes no se mostraron, el menú no funcionó bien. Estos fueron los problemas que se repitieron más a menudo.

Los principales resultados, después de la aplicación de la FG, fueron con respecto a la experiencia de los usuarios con el funcionamiento de la aplicación móvil. De acuerdo con esto, quince (83%) usuarios no tuvieron problemas al usar la aplicación y tres (17%) casos tuvieron una o tres interrupciones sin detener el proceso de evaluación.

Además, la Figura 5 muestra los resultados obtenidos de la aplicación de NASA-TLX, donde se puede mostrar que el promedio ponderado global para las dos tareas no supera el 50%, por lo tanto, realizar estas tareas tiene un factor de riesgo psicosocial equivalente de Bajo - Medio de la carga y cansancio mental.

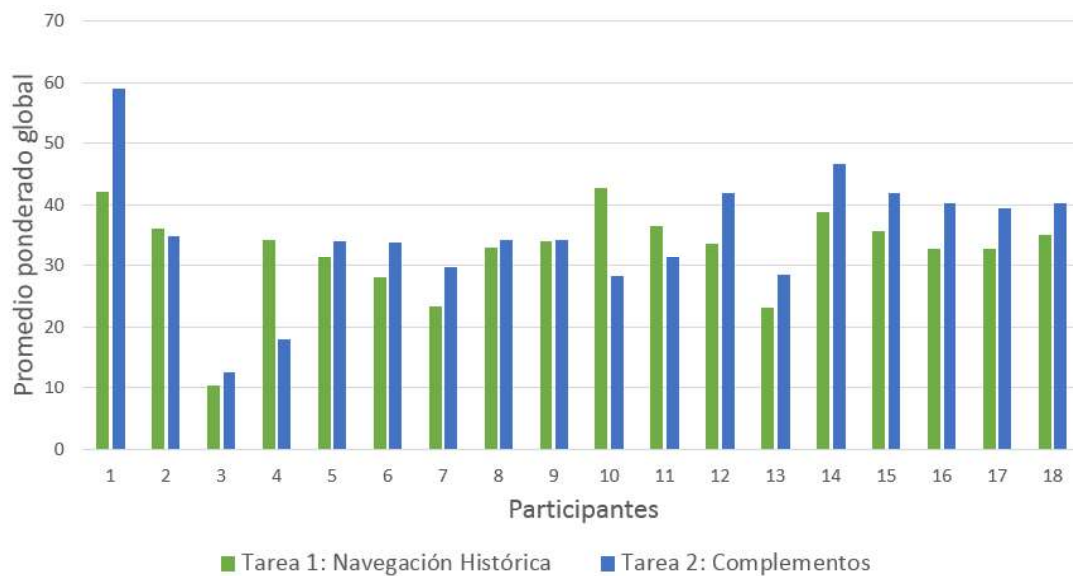


Figura 3: Promedio ponderado global -NASA - TLX - Usuarios del grupo experimental

Con cada uno de los datos obtenidos de las variables analizadas se puede observar en la Figura 6 que para ambas tareas la demanda mental, física y temporal es muy baja, ya que no es necesario memorizar ningún procedimiento, hacer ningún esfuerzo físico y se pausa y sin presión el ritmo de ejecución de la tarea. Además, los resultados muestran que los participantes alcanzaron los objetivos establecidos para la evaluación en la tarea 1 con un poco de esfuerzo. Sin embargo, para la tarea 2 el esfuerzo se incrementó y los objetivos establecidos no se lograron. Este último resultado es consistente con las pruebas realizadas con TA donde fue evidente que los usuarios tenían más errores de usabilidad relacionados con problemas de comprensión.

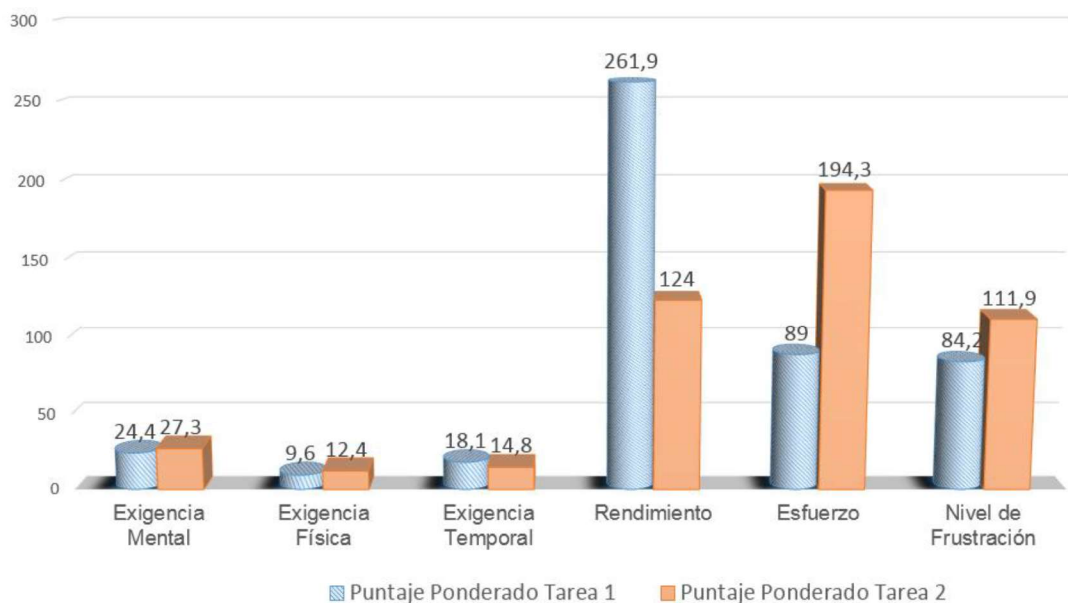


Figura 4: Puntaje ponderado de la NASA-TLX - Usuarios del grupo experimental

Por otro lado, la evaluación de usabilidad aplicada al grupo de control muestra otros resultados. El grupo de control tiene conocimiento del modo de operación para los actos de patrulla y delitos, ya que los cinco participantes son agentes de policía que han trabajado en el centro norte de la ciudad y en los valles de Quito, respectivamente. Los participantes incluyeron cuatro mujeres (80%) y un hombre (20%) con rangos de edad de entre 28 y 35 años. Se encontraron treinta errores con los métodos TA y FG. Sin embargo, los errores fueron reportados varias veces por diferentes usuarios.

La mayor parte de los errores encontrados en TA estaban relacionados con términos incorrectos o no intuitivos, la falta de ayudas textuales y errores de interfaz. En FG, la mayor parte de los errores fue de funcionalidad, específicamente con la información

que se muestra sobre los actos delictivos. Para resumir, se encontraron treinta errores informados por cinco usuarios, pero después del análisis y la categorización, se resumieron como doce errores. La Tabla 8 muestra la clasificación y la severidad de los errores que los usuarios encontraron en el proceso de evaluación.

Tabla 10: Clasificación y severidad de los problemas de usabilidad - Grupo de Control

Ítem	Problemas	Categoría	Severidad	Método
1	Términos no intuitivos	Problemas de terminología	2	TA
2	Botones sin información / etiquetas	Problemas de Comprensión	2	TA
3	Zoom en los mapas no funcionan	Problema de entrada de datos	3	TA/FG
4	Botones sin funcionalidad	Problemas de Comprensión	3	TA
5	Patrullaje no funciona	Problemas de Comprensión	3	TA/FG
6	Mensajes no se muestran	Problema de entrada de datos	2	FG
7	Botones sin ayuda	Problemas de Comprensión	2	FG
8	El menú no funciona correctamente	Problema de entrada de datos	2	TA/FG
9	Mapas sin leyendas	Problemas de terminología	2	TA
10	Información del delincuente no se actualiza	Problemas de Comprensión	3	FG
11	Información del sector no se actualiza	Problemas de Comprensión	3	FG
12	Información de la víctima no se actualiza.	Problemas de Comprensión	3	FG

De acuerdo con la tabla, los problemas de comprensión fueron los más frecuentes con siete (58%) de los casos. Alrededor (71%) de los problemas en esta categoría tenían una severidad de 3. En esta categoría, los principales problemas reportados fueron funciones como el patrullaje que no funciona, íconos no descriptivos, funciones cuyo propósito era desconocido, información no actualizada.

Los usuarios también identificaron tres (25%) problemas de entrada de datos. En esta categoría, tres problemas se clasificaron como severidad 3, el funcionamiento

incorrecto del zoom en los mapas, no se muestra ningún tipo de mensaje, el menú no funcionó bien. Estos fueron los problemas que se repitieron más a menudo.

Los principales resultados, después de la aplicación de la FG, fueron con respecto a la experiencia de los usuarios con el funcionamiento de la aplicación móvil. De acuerdo con esto, los usuarios no tuvieron problemas al usar la aplicación, pero el error principal estaba relacionado con la información en la aplicación móvil. Cada módulo debe actualizar la información de los delincuentes y las áreas con más actos delictivos. Además, los refuerzos más cercanos deben mostrarse en la opción de patrullaje para atender la emergencia cada vez más rápido.

Además, la Figura 6 muestra los resultados obtenidos de la aplicación de NASA-TLX con usuarios de control. De la misma manera que los usuarios experimentales, el promedio ponderado global para las dos tareas no excede el 50%, por lo tanto, realizar estas tareas tiene un factor de riesgo psicosocial Bajo-Medio equivalente a la carga y la fatiga mental.

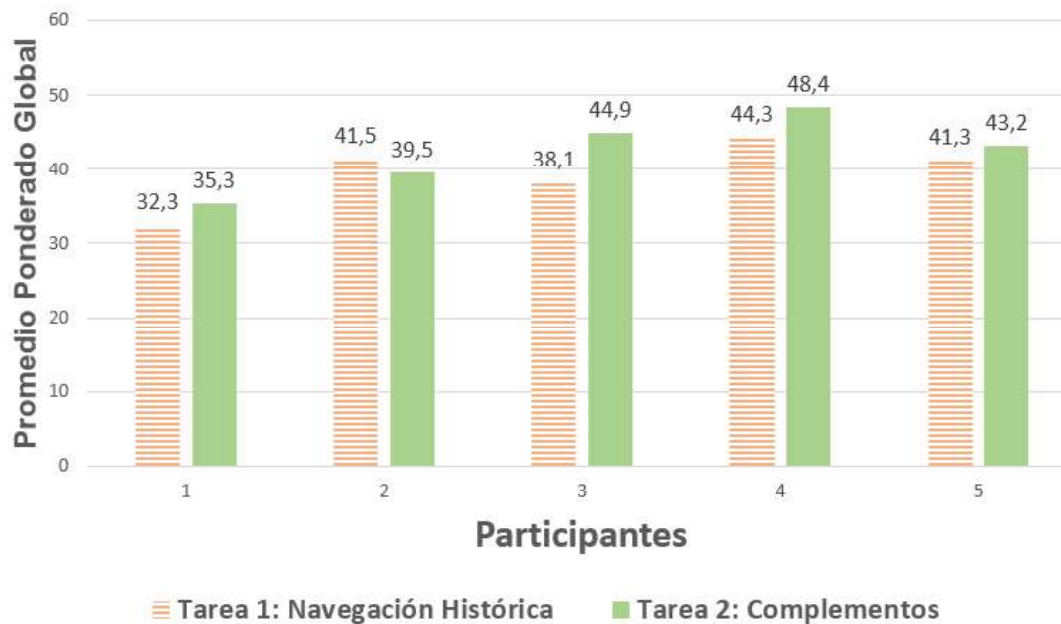


Figura 5: Promedio ponderado global NASA - TLX - Usuarios del grupo de Control

Con cada uno de los datos obtenidos de las variables analizadas, se puede observar en la Figura 7 que para ambas tareas el nivel físico, temporal y de frustración es muy bajo, ya que no es necesario memorizar ningún procedimiento ni realizar ningún

esfuerzo físico. El ritmo de ejecución de la tarea es pausado y sin presión. Además, los resultados muestran que los participantes alcanzaron los objetivos establecidos para la evaluación en la tarea 1 y con un poco más de esfuerzo para la tarea 2. La razón de esto fue la experiencia y el conocimiento del usuario sobre la lógica de la aplicación.

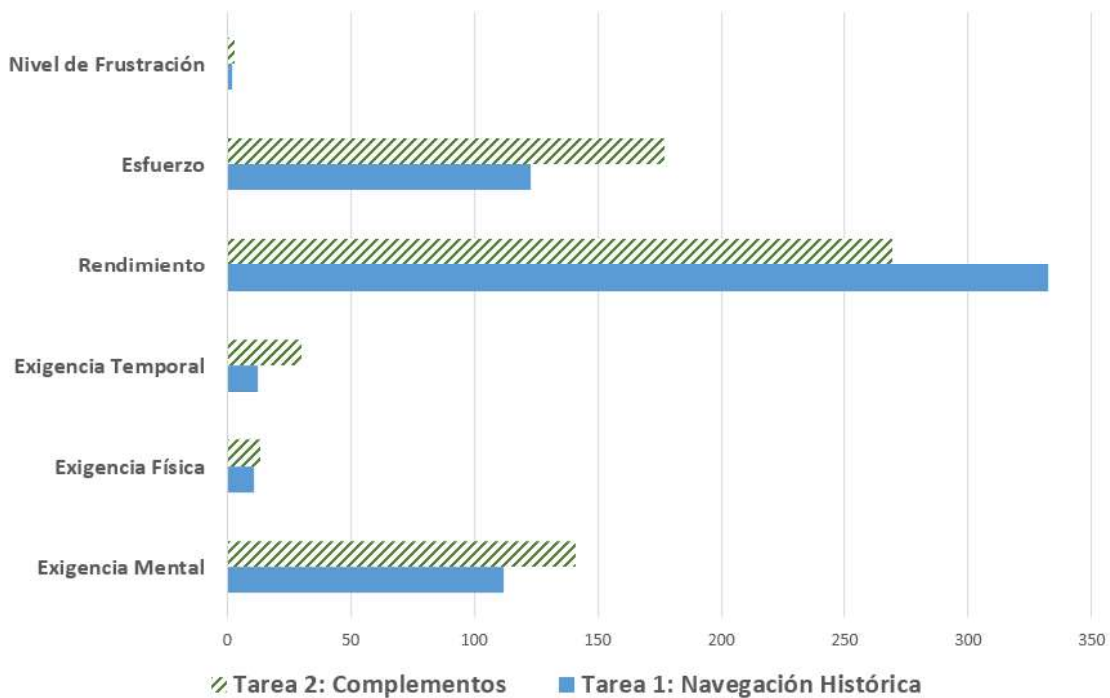


Figura 6: Puntaje ponderado de la NASA-TLX - Usuarios del grupo de Control

3.2. Reporte Técnico

Como resultado de esta investigación, se preparó un informe técnico con las mejoras propuestas para garantizar un mayor nivel de usabilidad de la aplicación móvil I-Pat. La tabla 8 muestra los errores encontrados con la mejora técnica respectiva. Las mejoras se priorizaron considerando los criterios de funcionalidad como prioridad 1, con el objetivo de que la aplicación móvil sea completamente funcional, evitando el problema de que el usuario no pueda realizar sus tareas. La prioridad 2 consideró los errores relacionados con la claridad y la facilidad de uso de la aplicación móvil, es decir, la implementación de ayudas para el usuario, términos intuitivos, iconos descriptivos, mensajes de ayuda y validación, evitando así la confusión para el

usuario. Finalmente, se consideraron los errores relacionados con las interfaces, que son errores que no detienen el funcionamiento de la aplicación, sin embargo, deben considerarse para que el usuario se sienta cómodo con el uso de la aplicación móvil.

La figura 7 muestra una propuesta para la interfaz de patrullaje de la aplicación. Se sugiere la aplicación de tres modificaciones: permitir el zoom en el mapa, reinicializar el mapa de acuerdo con la posición del usuario y reubicar la parte inferior de la sesión cerrada.

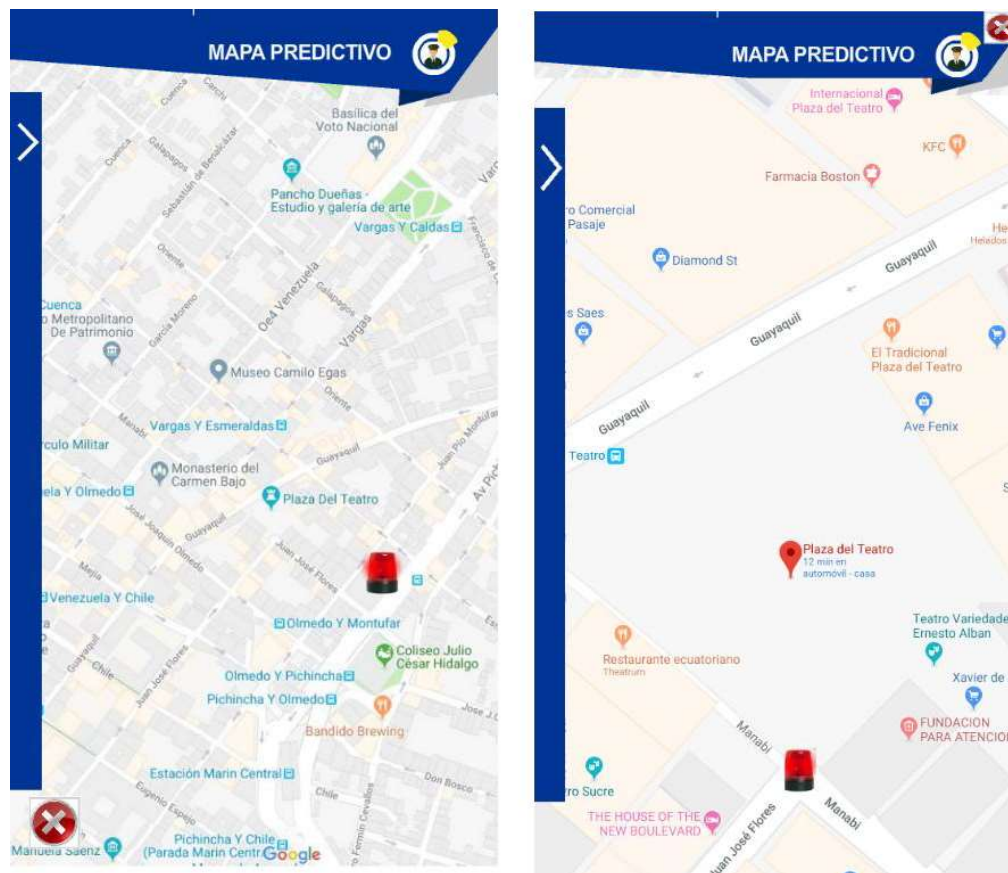


Figura 7: Interfaz de la aplicación I-Pat antes y después de la mejora.

Además, se recomienda llevar a cabo una capacitación previa sobre las funcionalidades de la aplicación móvil para los usuarios que interactúan con la aplicación, a fin de reducir el nivel de frustración y aumentar el nivel de rendimiento obtenido en la evaluación de NASA-TLX.

Tabla 11: Reporte Técnico con las recomendaciones para mejorar la usabilidad

N°	Problemas	Mejoras
1	Zoom en el mapa no funciona	Comprobar la integración de la API de Google en la aplicación, también habilitar las funciones de zoom dentro de los mapas.
2	Información del delincuente no se actualiza	Verificar que la información se actualice de acuerdo a la funcionalidad indicada.
3	Información del sector no se actualiza	Verificar que la información se actualice de acuerdo a la funcionalidad indicada.
4	Información de las víctimas no se actualiza	Verificar que la información se actualice de acuerdo a la funcionalidad indicada.
5	Patrullaje no funciona	Revise la implementación del algoritmo k-means para realizar correctamente el patrullaje
6	Botones sin funcionalidades	Asigna los métodos correctamente a cada botón dentro de la aplicación caso contrario eliminar el botón.
7	El menú no trabaja bien	Verifique que las clases y los métodos asignados al menú realicen las tareas de acuerdo con los requisitos. Además, debe verificar que la tarea correspondiente esté habilitada de acuerdo al perfil de usuario.
8	Mensajes no se muestran	Verificar que los mensajes de confirmación y validación estén implementados para facilitar las tareas realizadas por el usuario dentro de la aplicación.
9	Funciones no intuitivas	Implementar mensajes en cada una de las funciones detallando su principal objetivo.
10	Mapas sin ayuda	Los mapas deben incluir ayudas que permitan ingresar la ubicación actual, el destino y el tiempo que tomará para una ruta determinada. Además debe permitir la visualización de rutas alternativas.
11	Nombres en los botones poco intuitivos	Los nombres en los botones deben contener nombres descriptivos, que permitan al usuario entender cuál es el propósito de esa función. O, a su vez, debe incluirse un enlace al manual del usuario dentro de la aplicación.
12	Términos no intuitivos	Utilizar terminología que el usuario pueda entender. Se recomienda cambiar las etiquetas a términos más descriptivos.

13	No hay ayuda para el usuario	Se recomienda tener un manual del usuario, colocar el texto de la sugerencia de la herramienta en los botones, mostrando la información de contacto para ayudar a los usuarios.
14	Iconos sin descripción	Este problema está relacionado con los puntos 9, 11 y 12. También se recomienda que los botones, menús e imágenes contengan una etiqueta con un nombre descriptivo.
15	Botones mal ubicados	Los botones deben estar ubicados correctamente, sin bloquear el acceso a otras funciones dentro de la aplicación.
16	Colores no adecuados	Se recomienda seguir algunos principios básicos para el manejo de colores como la teoría, los modelos, los esquemas de colores que permiten, a través de un estudio, determinar los colores que ayudan al usuario a sentirse cómodo con la aplicación.
17	Distribución del menú incorrecta	Se recomienda priorizar el orden de las funciones en el menú, debe colocar las funciones con las que trabaja continuamente hasta que alcance el nivel más bajo con funciones poco utilizadas.
18	No se permite la personalización	Se recomienda implementar una personalización de aspectos como: colores, tonos, tipos de letra e idioma con los que el usuario se sienta más cómodo y pueda acceder más fácilmente a la aplicación.

3.3. Discusión

Hoy en día, el uso de tecnologías móviles dentro de la administración gubernamental para prestar servicios públicos a ciudadanos, funcionarios y empresas ha ido en aumento. El gobierno de Ecuador, priorizando la seguridad como una necesidad primaria de los ciudadanos, ha centrado sus esfuerzos en apoyar el trabajo de la Policía Nacional, reforzando el servicio de seguridad ciudadana con nuevas tecnologías. Inicialmente enfocaron su desarrollo una aplicación web. Sin embargo, tras analizar las ventajas de disponer una aplicación móvil que facilite y agilice el acceso a la información en tiempo real, permitiendo al oficial de policía no entorpecer su trabajo, optimizando los recursos de sus dispositivos móviles como la cámara de video para grabar a los delincuentes

cometiendo los crímenes, reportando su ubicación actual de manera ágil permitiendo que los refuerzos lleguen de una manera más rápida, compartiendo información del delincuente al instante, enviando información del circuito donde se realiza la persecución. Con estas ventajas, el Ministerio del Interior desarrolló la aplicación móvil I-Pat con el propósito de proporcionar una herramienta para apoyar a los oficiales de policía en sus tareas de patrullaje, administrando información extensa sobre circuitos criminales, crímenes e información de criminales. La evaluación de la usabilidad de la aplicación móvil I-Pat tiene como objetivo minimizar el impacto en los posibles problemas de usabilidad y, por lo tanto, permitir un mejor uso por parte de los usuarios finales. Por esta razón, es importante tener un estudio de evaluación de usabilidad que motive al oficial de policía a usar la aplicación móvil con seguridad, eficacia, eficiencia y satisfacción.

Este estudio integró dos enfoques diferentes de evaluación de usabilidad y NASA-TLX para la carga de trabajo. Los dos tipos diferentes de análisis de usabilidad nos dieron una comprensión multifacética de lo que los usuarios esperan de las tareas de la aplicación y cómo esperan interactuar con la aplicación. Según la literatura, la revisión sugiere el uso de TA para la evaluación de la usabilidad [50], [35]. Por otro lado, en [37] propone el uso de FG como una técnica de investigación cualitativa ampliamente difundida en varios campos de investigación en psicología y otras disciplinas científicas.

La presente investigación combina las ventajas de ambos métodos para encontrar un mayor número de problemas de usabilidad. Como puede ver en los resultados, hubo problemas que se informaron con ambos métodos y otros que correspondían solo a TA y FG respectivamente (ver Tablas 6, 7 y 8). Debido a la naturaleza de cada método, los resultados se verán influenciados por cómo se siente el usuario cuando expresa sus pensamientos y conclusiones. Algunos prefieren internalizar y analizar su experiencia antes de usar la aplicación móvil y luego expresarla y discutirla con más usuarios, mientras que otros tienen la capacidad de comunicar instantáneamente lo que piensan al usar y probar la aplicación móvil.

Además, las clasificaciones subjetivas de la carga de trabajo mental se midieron mediante NASA-TLX. En la investigación [41], los autores mencionan la importancia de la carga cognitiva como un atributo de usabilidad. En ese estudio se encontró que casi el 23% de los estudios midió la carga cognitiva de la aplicación bajo evaluación. La puntuación de la NASA-TLX varía de 0 (sin esfuerzo) a 100 (esfuerzo máximo) [10]. Con la aplicación de NASA-TLX fue posible demostrar que la carga de trabajo de la aplicación no supera el 50%. Sin embargo, para lograr el máximo rendimiento de la aplicación, se recomienda realizar una capacitación previa antes de usarla, de modo que los usuarios puedan explorar las opciones sin dificultad y sin un alto nivel de frustración, como fue el caso del grupo de usuarios experimentales.

Con la aplicación de TA, FG y NASA, TLX se proporcionó a un equipo de desarrolladores de la Policía Nacional de Ecuador información relevante, un método de usabilidad para optimizar la implementación de la aplicación y una predicción de la carga de trabajo cognitiva. En este sentido, hicimos sugerencias para resolver los problemas de usabilidad y mejorar el rendimiento de la aplicación. La mayoría de estas sugerencias fueron sobre la solución de problemas funcionales, un mejor diseño de algunos botones, además de incluir ayudas o textos descriptivos en los iconos de las diferentes funciones y realizar una capacitación para usuarios que trabajan diariamente con la aplicación.

Debido al acceso restringido a la aplicación, no fue factible realizar otra evaluación de usabilidad después de la implementación de las recomendaciones técnicas realizadas. Sin embargo, debido a la importancia de la aplicación para mejorar el servicio de patrullaje y el compromiso de la Institución para proporcionar a sus empleados una aplicación que respalde sus tareas diarias, se resolverán los problemas funcionales y se darán conferencias sobre el uso de la aplicación.

4. CONCLUSIONES

La seguridad es una necesidad importante en el desarrollo de la sociedad. Por esta razón, la Policía Nacional de Ecuador buscó una solución llamada I-Pat que, con el apoyo de la tecnología, permite patrullar las áreas con más conflictos, proporcionando así un grado de confort a la población. Al ser un sistema de tanta importancia, la evaluación de la usabilidad permite al usuario utilizar el sistema con mayor facilidad, familiarizándose más con las funciones y apoyando así el trabajo de la Policía Nacional.

Las pruebas de usabilidad por sí mismas no pueden desarrollar una lista completa de defectos, para esto, es importante utilizar una combinación adecuada de métodos. En este contexto, esta investigación propuso una nueva forma de evaluar la usabilidad mediante la combinación de TA y FG como métodos de usabilidad y NASA-TLX para la carga de trabajo cognitiva. Estos métodos que combinan sus ventajas permitieron cubrir un mayor número de problemas dentro de la aplicación.

Este estudio mostró que una gran cantidad de los problemas identificados durante el proceso de evaluación se relacionaron con problemas de comprensión, esto se debió al diseño deficiente del sistema de acuerdo con las necesidades de los usuarios. Por lo tanto, se recomienda que los usuarios sean educados con las funciones de la aplicación antes de utilizarla.

La evaluación de la usabilidad en las primeras etapas del ciclo de vida del software traerá grandes ventajas, ya que los programas de software, las webs o los productos de consumo pueden diseñarse con mayor precisión, sin anticipar intuitivamente las acciones del cliente, sino predecirlas con bastante precisión.

Como trabajo futuro, se realizará un estudio con una nueva versión de la aplicación, que incluye las mejoras propuestas en el informe técnico, también podría incluirse el enfoque de expertos en usabilidad.

5. REFERENCIAS BIBLIOGRÁFICAS

- [1] International Organization for Standardization ISO 25010, Systems and software engineering. Available online: <https://www.iso.org/standard/35733.html>. (2011).
- [2] Jiang, Y., Wu, D., Liu, P. (2016) JRed: Program Customization and Bloatware Mitigation Based on Static Analysis. In Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (1), 12–21. <https://doi.org/10.1109/compsac.2016.146>
- [3] Brhel, M., Meth, H., Maedche, A., & Werder, K. (2015). Exploring principles of user-centered agile software development: A literature review. *Information and software technology*, 61, 163-181.
- [4] Ivanc, D., Onita, M. (2015) Usability evaluation of a LMS mobile web interface. In Proceedings of the International Conference on Information and Software Technologies, 348–361. https://doi.org/10.1007/978-3-642-33308-8_29
- [5] Ryu, Y. S. (2005). Development of usability questionnaires for electronic mobile products and decision making methods (Doctoral dissertation, Virginia Tech).
- [6] Coursaris, C., & Kim, D. (2006). A qualitative review of empirical mobile usability studies. *AMCIS 2006 Proceedings*, 352.
- [7] Biel, B., Grill, T., & Gruhn, V. (2010). Exploring the benefits of the combination of a software architecture analysis and a usability evaluation of a mobile application. *Journal of Systems and Software*, 83(11), 2031-2044
- [8] Wesson, J. L., Singh, A., & Van Tonder, B. (2010, September). Can adaptive interfaces improve the usability of mobile applications? In *IFIP Human-Computer Interaction Symposium* (pp. 187-198). Springer, Berlin, Heidelberg.
- [9] Guevara, C., Jadán, J., Zapata, C., Martínez, L., Pozo, J., & Manjarres, E. (2018). Model of Dynamic Routes for Intelligent Police Patrolling. In *Multidisciplinary Digital Publishing Institute Proceedings* (Vol. 2, No. 19, p. 1214). Proceedings of UCAml 2018, <https://doi.org/10.3390/proceedings2191214>.
- [10] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

- [11] Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006, April). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 1253-1262). ACM.
- [12] Sánchez, W. O. (2015). La usabilidad en Ingeniería de Software: definición y características [Usability in software engineering: definition and characteristics]
- [13] Caplan, S. (1990). Using focus group methodology for ergonomic design. *Ergonomics*, 33(5), 527-533.
- [14] Nielsen, J. (1997). The use and misuse of focus groups. *IEEE software*, 14(1), 94-95.
- [15] Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior research methods*, 41(1), 113-117.
- [16] Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.
- [17] Park, S., Harada, A., & Igarashi, H. (2006, April). Influences of personal preference on product usability. In CHI'06 Extended Abstracts on Human Factors in Computing Systems (pp. 87-92). ACM.
- [18] Kitchenham, B. A. (2012, September). Systematic review in software engineering: where we are and where we should be going. In Proceedings of the 2nd international workshop on Evidential assessment of software technologies (pp. 1-2). ACM
- [19] Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010, April). Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2381-2390). ACM.
- [20] McDonald, S., McGarry, K., & Willis, L. M. (2013, September). Thinking-aloud about web navigation: the relationship between think-aloud instructions, task difficulty and performance. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 57, No. 1, pp. 2037-2041). Sage CA: Los Angeles, CA: SAGE Publications.
- [21] Hertzum, M., Hansen, K. D., & Andersen, H. H. (2009). Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28(2), 165-181.
- [22] Barnum, C. M. (2010). *Usability testing essentials: ready, set... test!* Elsevier.

- [23] Van Den Haak, M. J., De Jong, M. D., & Schellens, P. J. (2006). Constructive interaction: An analysis of verbal interaction in a usability setting. *IEEE transactions on professional communication*, 49(4), 311-324.
- [24] Makri, S., Blandford, A., & Cox, A. L. (2011). This is what I'm doing and why: Methodological reflections on a naturalistic think-aloud study of interactive information behaviour. *Information Processing & Management*, 47(3), 336-348.
- [25] Nørgaard, M., & Hornbæk, K. (2006, June). What do usability evaluators do in practice? an explorative study of think-aloud testing. In *Proceedings of the 6th conference on Designing Interactive systems* (pp. 209-218). ACM.
- [26] Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004). Comparative usability evaluation. *Behaviour & Information Technology*, 23(1), 65-74.
- [27] Gulliksen, J., Boivie, I., Persson, J., Hektor, A., & Herulf, L. (2004, October). Making a difference: a survey of the usability profession in Sweden. In *Proceedings of the third Nordic conference on Human-computer interaction* (pp. 207-215). ACM.
- [28] O Nyumba, T., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, 9(1), 20-32.
- [29] Plummer-D'Amato, P. (2008). Focus group methodology Part 2: Considerations for analysis. *International Journal of Therapy and Rehabilitation*, 15(3), 123-129.
- [30] Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A qualitative framework for collecting and analyzing data in focus group research. *International journal of qualitative methods*, 8(3), 1-21.
- [31] Kontio, J., Lehtola, L., & Bragge, J. (2004, August). Using the focus group method in software engineering: obtaining practitioner and user experiences. In *Proceedings. 2004 International Symposium on Empirical Software Engineering, 2004. ISESE'04.* (pp. 271-280). IEEE.
- [32] Sunikka, A. (2004). Usability evaluation of the Helsinki School of Economics Website. Master's thesis.
- [33] Obeso, A., & Elena, M. (2004). Metodología de medición y evaluación de la usabilidad en sitios web educativos.
- [34] Nielsen, J. (2012). Why you only need to test with 5 users, 2000. Jakob Nielsen's Alertbox. Available on line from [www. useit. com/alertbox/20000319. html](http://www.useit.com/alertbox/20000319.html).

- [35] Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: a practical approach to modelling cognitive*. Academic Press, London.
- [36] Templeton, J. F. (1994). *The focus group: a strategic guide to organizing, conducting and analyzing the focus group interview*. Probus Publishing Company.
- [37] Wilkinson, S., & Silverman, D. (2004). 10 Focus Group Research. *Qualitative research: Theory, method and practice*, 177-199.
- [38] Harper, B. D., & Norman, K. L. (1993, February). Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5. In *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference* (pp. 224-228).
- [39] Arh, T., & Blazic, B. J. (2008). A case study of usability testing—the SUMI evaluation approach of the EducaNext portal. *WSEAS transactions on information science and applications*, 5(2), 175-181.
- [40] Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- [41] Angulo, J. S., & Robles, D. C. (2014). Usabilidad y satisfacción de la e-Rúbrica. *REDU. Revista de Docencia Universitaria*, 12(1), 177-195. *REDU. Revista de Docencia Universitaria*, 12(1):177–195, 2014.
- [42] Tracy, J. P., & Albers, M. J. (2006, May). Measuring cognitive load to test the usability of web sites. In *Annual Conference-society for technical communication* (Vol. 53, p. 256).
- [43] Pilco, H., Sanchez-Gordon, S., Calle-Jimenez, T., Pérez-Medina, J. L., Rybarczyk, Y., Jadán-Guerrero, J., ... & Nunes, I. L. (2019). An Agile Approach to Improve the Usability of a Physical Telerehabilitation Platform. *Applied Sciences*, 9(3), 480
- [44] Habibi, M. R. M., Khajouei, R., Eslami, S., Jangi, M., Ghalibaf, A. K., & Zangouei, S. (2018). Usability testing of bed information management system: A think-aloud method. *Journal of advanced pharmaceutical technology & research*, 9(4), 153.
- [45] Tang, K. C., & Davis, A. (1995). Critical factors in the determination of focus group size. *Family practice*, 12(4), 474-475.
- [46] Calle-Jimenez, T., Tutillo, D., Sanchez-Gordon, S., Jadán-Guerrero, J., Guevara, C., Lara-Alvarez, P., Acosta-Vargas, P., Salvador-Ulluari, L., Nunez, I., (2019). Improving Usability with Think Aloud and Focus Group Methods. Case Study: An Intelligent Police Patrolling System (I-Pat).

- [47] Van den Haak, M. J., de Jong, M. D., & Schellens, P. J. (2004). Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: a methodological comparison. *Interacting with computers*, 16(6), 1153-1170.
- [48] Nielsen, J. (1995). Severity ratings for usability problems. *Papers and Essays*, 54, 1-2.
- [49] Tutillo, D., Calle-Jimenez, T., Sanchez-Gordon, S., (2019). Assessing the Usability of Intelligent Police Patrolling Mobile Application.
- [50] Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science*, 1(1), 1.

6. ANEXOS

6.1. Artículo resultado de la primera iteración

6.2. Artículo resultado de la segunda iteración

6.3. Plantilla del cuestionario de usabilidad

6.4. Plantilla del cuestionario NAS_TLX

Improving Usability with Think Aloud and Focus Group Methods. A Case Study: An Intelligent Police Patrolling System (I-Pat)

Tania Calle-Jimenez¹, Doris Tutillo-Sanchez¹, Sandra Sanchez-Gordon¹, Janio Jadán-Guerrero², Cesar Guevara², Patricio Lara-Alvarez², Patricia Acosta-Vargas³, Luis Salvador-Ullauri⁴, Isabel L. Nunes⁵

¹ Escuela Politécnica Nacional, Quito, Ecuador

² Universidad Tecnológica Indoamérica, Quito, Ecuador

³ Universidad de Las Américas, Quito, Ecuador

⁴ Centro de Educación Continua, Quito Ecuador

⁵ Universidad de Nova de Lisboa, Lisboa-Portugal

{`tania.calle`, `doris.tutillo`, `sandra.sanchez`}@epn.edu.ec
{`janiojadan`, `cesarguevara`, `patolara`}@uti.edu.ec `patricia.acosta@udla.edu.ec`,
`lsalvador@cec-epn.edu.ec`, `imm@fct.unl.pt`

Abstract. This study proposes the use of Think Aloud and Focus Group methods for evaluating the usability of the Intelligent Police Patrolling System (I-Pat). The study was conducted with twenty-one students of computer engineering. The study included two evaluation methods. First, the application of Think Aloud using audio recordings, image capturing and questionnaires. Second, the application of a Focus Group for brainstorming. The total number of the usability problems identified was fifteen. Comprehensiveness (46%) and layout (43%) problems were the most frequently found. The study showed that the problems encountered were due to the lack of understanding of the system's functions, so it is recommended increasing the users' levels of knowledge about the system. The application of these methods caused the students to find a greater number of errors than when applying a single method, allowing them to generate a report with usability improvements according to the reported errors.

Keywords: Think Aloud · Focus Group · Usability · Patrolling · I-Pat.

1 Introduction

The International Organization for Standardization in ISO/IEC 9126-1 defines usability as the ability of the software product to be understood, learned, used and be attractive to the user, when used under specific conditions [1]. However, this concept does not receive the appropriate importance, which implies confusion, error, delay or absolute failure to complete some task by the user [2].

Nowadays, this is changing, promoting an evaluation of the initial phases of the software development process and a more active participation by the user in the later phases

of the cycle, including usability that is considered a crucial non-functional requirement for economic success in competitive markets [3], [4].

This research evaluates the usability of the Intelligent System of Police Patrols (I-Pat) developed by the Ministry of the Interior in Ecuador. This system covers one of the fundamental needs for the Ecuadorian community, that is security with technology, patrols and surveillance of cities to identify crimes and criminal groups. I-Pat is a model that generates dynamic routes applying artificial intelligence through the K-means algorithm to identify critical points when patrolling a circuit within the Ecuadorian territory. In addition, I-Pat has incorporated the Google Maps API tool to design routes and verify means of transport. The usability evaluation of I-Pat allowed the measurement of the capacity of the system to be understood, learned and used by police officers. The Think Aloud (TA) and Focus Group (FG) evaluation methods were used to propose an evaluation that emphasizes aspects such as the global structure of the site, focusing on content and navigability, designed to offer clear paths from the origin to the destination, without complex terminology and with the mission of guiding the user. Finally, the consistency of the contents, as well as the visual format, was evaluated. The disposition of elements in the pages offer homogeneous environments that help to promote an effective communication of the message. At the same time, the consistency helps the user to form a mental model of the site [5].

TA allows collection of direct verbal information from the user about their thoughts. This includes cognitive processes and strategies while participating in a task. This method ranges from open and unstructured questions to structured interviews that can be adjusted according to the responses and behavior of the user [6]. In addition, TA is a technique used to understand cognitive processes based on tasks requested from users, both within usability tests and in the broader study of human-computer interaction. In this study, the evaluation was applied to twenty-one users; users interacted with the system's interfaces and described the actions and tasks they performed. Users reported inconsistencies, confusion, concepts not understood, unfinished actions, as well as those aspects that seemed good about the interaction.

FG is a focused discussion in which a moderator addresses a group of participants through a set of questions about a particular topic. This method consists of a planned and designed discussion about the interfaces. In this research, the discussion was guided and facilitated by a moderator. The moderator followed a predefined structure so that the discussion remained focused. Two groups were formed and participated in separate discussion sessions. The activity that was used was brainstorming, in order for the participants to exchange opinions. This research presented the results of the usability evaluation in a technical report that contains recommendations to improve the usability level of the I-Pat system.

2 Literature Review

In the literature review, authors have identified the use of TA and FG as methods to measure usability of various software products, [7] defines TA as a usability evaluation method used to gain insight into how people work with a product or interface, [8] argues that usability professionals currently use variations of TA as the main method to identify usability problems. This study compares three TA Protocols for using in testing. The

first protocol, called Instruction, is where there are no instructions for varying amounts of instruction or practice sessions, then Intervention: different types of intervention with a variety of probes or prompts, and Prompting: varying rates of the test with the administrators administering a prompt. This study shows variations of TA with these protocols that usability practitioners currently use, the practitioners who want measures that reflect unaided user behavior can choose between two methods traditional or speech communication that do not differ significantly from each other in terms of their effects on user performance. In the [9] study the concurrent think-aloud method was used. This is a highly valued technique that is used to understand peoples' task-based cognitive processes, within both usability testing and the broader study of human-computer interaction. In addition, [10] showed that a relaxed think-aloud, which included evaluator interventions, increased test time and led to changes in behavior at the interface, [11] has other approaches to gathering verbal data; these include constructive interaction, where users work together in pairs. Constructive interaction can take a number of forms: users can work together on an equal footing [12] or alternative formats may be used such as teach-back scenarios, in which the user teaches another person about the focus system. In [13] comments users said that they may not give reasons for their actions during think-aloud, only a description of the actions themselves, [14] proposes some tools to capture the observations during the sessions using TA. Other studies have focused more on tactical questions of usability evaluation, for example [15] presents an example of a results report that should be presented after the evaluation with TA.

The literature suggests that TA is a widely used method for evaluating usability by taking its initial version proposed by Ericson and Simon or with different versions obtaining good results, as shown by [16] where it was discovered that think-alouds were classified among the five main methods used by the usability professionals in Sweden over prototyping, interviews, field studies and scenarios.

On the other hand, FG have been used for decades as mentioned in [17]. Discussion is frequently used as a qualitative approach to gain an in-depth understanding of social issues. The method aims to obtain data from a purposely selected group of individuals rather than from a statistically representative sample of the broader population. FG are less threatening to many research participants, and this setting is useful for participants to discuss perceptions, ideas, opinions and thoughts. For example, [18] proposes a series of steps to maximize the reliability of the data and recommends some criteria like reliability, internal validity, transferability and confirmability to ensure that important aspects of qualitative research are reported, which increases the quality of the published studies of FG.

[19] specifies a framework to capture and analyze the information derived from FG in relation to the type of information generated, quasi-statistical methods and conversation analysis. In addition, [20] proposes that FG can provide a valuable and complementary empirical experience quickly and at low cost, [21] uses the method to identify usability characteristics with respect to an institutional website.

However, as mentioned by [22], although focus groups can be a powerful tool in software development, it should not be used as the only source of usability data, [23] reviews some of the characteristics of the techniques for usability and considers TA as a complement in the application of FG. Consequently, the present research will use TA and FG as usability evaluation methods for the I-Pat system.

3 Methods

When considering safety as a global society need, this study applies the TA and FG evaluation methods to the police patrol system developed by the National Police of Ecuador. One of the objectives of the application of these methods is to obtain usability errors according to the characteristics of their participants, such as their behavior, gender, age and knowledge of mobile systems. The I-Pat system, the participants in the evaluation and the evaluation process are described in detail below.

3.1 I-Pat System

I-Pat is the acronym for Intelligent Police Patrolling. This system covers one of the fundamental needs for the Ecuadorian community, i.e. safety. With the use of technology, patrols and surveillance of cities will be facilitated by identifying crimes and criminal groups, helping the work of the national police. In this context I-Pat was developed for applications that use the Android system in order to perform intelligent patrolling for the National Police of Ecuador. For this purpose, it uses information about the crimes reported in three circuits during a period of four years [24], supported by a model that generates dynamic routes applying artificial intelligence with the K-means algorithm to identify critical points when patrolling a circuit within the Ecuadorian territory. In addition, the Google Maps API tool is used to design the routes and means of transport that the police officer will use. The I-Pat system has historical modules detailing the current information available regarding theft from people, theft from houses, theft from banks, theft of vehicles, theft of motorcycles, theft of accessories and theft on roads. It also has a complementary section with functions which include requests for police reinforcements patrols, predictive maps, schedules of the agents and civil unrest. Fig 1 shows a sample of interfaces of the patrol, civil unrest and vehicle theft modules.



Fig. 1. I-Pat System Interfaces

3.2 Participants' Test

In this research, the usability of the I-Pat System was evaluated with the participation of students from the Escuela Politécnica Nacional of Ecuador. Participants in this study consisted of two groups, including testers and facilitators. Testers included twenty-one computer engineering students in their sixth and seventh semester who are trained to do the quality assessment of different systems, since students have approved the software quality course. Testers had knowledge about designing and analyzing computer systems as

well as the principles and concepts of programming languages but no knowledge and experience using I-Pat. Before starting the evaluation, all the users were verbally familiar with the study's aims, and their informed consent was obtained. For the first TA assessment method, the facilitator reminded users to express their thoughts aloud. While, for the second FG method, the facilitator took the role of moderator, directing the groups to obtain the greatest number of ideas regarding the usability of I-Pat.

3.3 Evaluation Process

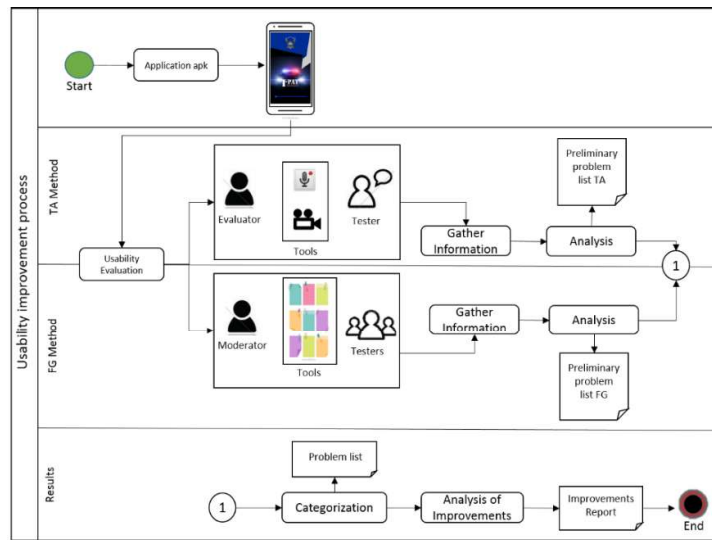


Fig. 2. Usability Evaluation

Fig 2 shows the process followed to apply TA and FG to evaluate the usability of the I-Pat system. To start with, the evaluation users were asked to install the application on their mobile phones with the Android operating system. Some testers used an emulator installed on a desktop machine, because they did not have a mobile with this operative system. Once testers had the application started, an introduction to the system's functions was done to familiarize the users.

The first session used TA, the method that allows the collection of information about the users' cognitive interaction with the system [25]. The users were asked to express what they saw, thought and felt with respect to the use of the I-Pat interfaces. After this introduction, the TA ran for 30 minutes and, during this session the facilitator reminded the users to express their thoughts aloud and audio recorded them for later analysis.

The second session worked with the FG method. Two groups were organized, the moderator brainstormed with them with the aim of obtaining original ideas about the evaluation of the usability of the system, through the free exchange of ideas between each of the testers. The moderator explained the rules of the session, such as avoiding criticism and expressing all the ideas that the users have. The moderator also mentioned that all thoughts were important even if those thoughts did not appear to be common

sense because those ideas could serve as inspiration for the development of new proposals. Testers expressed their ideas to the rest of the group.

3.4 Questionnaires

In addition, users filled out a questionnaire consisting of twenty questions to measure usability. This questionnaire was based on QUIS [26], SUMI [27] and SUS [28].

The questionnaire applied in this research consisted of four sections. The first contained questions about demographic details, such as age, gender and education. It also focused on the participants' experience in working with mobile devices. The second section was focused on learning about the system, and the third section worked on topics related to system capacity. The final section worked with usability and user interfaces. The items are presented on a scale from zero to nine, zero being confusing and nine being clear.

4 Results

Finally, the audio, images and questionnaire applied for TA and the brainstorming obtained with the FG were available for analysis. This study was conducted to analyze the usability problems of the I-Pat System using TA and FG methods. Participants included thirteen (62%) males and eight (38%) females with age ranges of between 18-22 (66.7%), 23-27 (28.6%) and 28-32 (4.8%). Eighteen errors were found with the TA method; however, errors were reported several times by different users.

Table 1. TA and FG results

Item	Problems	TA	FG
1	Sections without help	x	
2	Non-intuitive terms	x	x
3	Buttons without information/labels	x	x
4	Zoom in maps not working	x	x
5	Button close session is wrongly located	x	x
6	Buttons without functions	x	x
7	Patrol did not work	x	
8	Incorrect menu distribution	x	
9	Maps without help	x	
10	Personalization was not allowed		x
11	Messages were not displayed		x
12	Without user manual		x
13	Buttons without help	x	x
14	Unsuitable colors in interfaces		x
15	The menu did not work well		x

Table 1 shows the errors identified in the TA and FG sessions, the greater part of the errors found in TA were related to incorrect or unintuitive terms, the lack of textual aids

or a user manual. In addition, functionality errors were reported. On the other hand, thirty-four errors were reported with the FG method, with this method the errors of functionality and the lack of help found with the TA are maintained and interface errors are added. To summarize, fifty-two errors reported by twenty-one users were found but, after the analysis and categorization, these were summarized as fifteen errors. The errors found were classified according to the Haak method [12], this method establishes four main categories described in Table 2.

Table 2. Haak method categories

Item	Category	Description
1	Layout problems	The participant fails to spot a particular element on a screen.
2	Terminology problems	The participant does not comprehend part(s) of the terminology.
3	Data entry problem	The participant does not know how to interact with the search.
4	Comprehensiveness problems	The system lacks information necessary to use it effectively.

In addition, the Nielsen method [29] was used to identify the severity of the problems found according to the classification in Table 3.

Table 4 shows the classification and severity of the errors which users encountered in the assessment process. According to the table, comprehensiveness problems were the most frequent with seven (46%) of cases. Around 43% of the problems in this category had a severity of 2 and 3. In this category, the main problems reported were the malfunction of the zoom in the maps (3), functions such as the patrol not working (12), non-descriptive icons (18), functions whose purpose was unknown (6). Then six (40%) layout problems were identified by the users. In this category, four problems were rated as Severity 2, the problems related to lack of help in the icons (18), the absence of messages (6), inappropriate colors (5). These were the ones which recurred most often.

The main results, after the application of the questionnaire, were regarding the users' experience with the operation of the system. According to this, fifteen (71%) users had no problem using the application and four (19%) cases had one to three interruptions without stopping the evaluation process.

Table 3. Severity classification

Item	Problems
0	No usability problem
1	Cosmetic problem
2	Minor usability problem
3	Major usability problem
4	Usability catastrophe

Table 4. Errors found

Item	Problems	Category	Severity	Method
1	Zoom in maps not working	Data entry problem	3	TA/FG
2	Incorrect menu distribution	Layout problems	1	TA
3	Icons without description	Comprehensiveness problems	2	FG
4	Buttons without functions	Comprehensiveness problems	3	TA/FG
5	Titles on the buttons made no sense	Comprehensiveness problems	3	TA
6	Map without help	Comprehensiveness problems	3	TA
7	Misplaced buttons	Layout problems	1	TA/FG
8	Unsuitable colors	Layout problems	1	FG
9	The menu did not work well	Data entry problem	3	FG
10	Personalization was not allowed	Layout problems	1	FG
11	Messages were not displayed	Data entry problem	3	FG
12	Non-intuitive functions	Comprehensiveness problems	3	FG
13	There was no help for the user	Comprehensiveness problems	2	TA/FG
14	Patrol did not work	Comprehensiveness problems	3	TA
15	Non-intuitive terms	Terminology problems	2	TA/FG

Furthermore, two (9%) problems remained as unresolved after the user's review, and the evaluation process continued without that task. After the evaluations, 90% of the users made proposals for improvements to the system. Most of the suggestions were related to better design (57%) and improvements with the functions that Google maps use (21%).

The main results after the application of the questionnaire, it was regarding the users experience with the operation of the system. According this 15 (71%) users had no problem using the application and 4 (19%) cases had one to three interruptions without stopping the evaluation process. Furthermore, 2 (9%) problems remained as unresolved after the users review, and the evaluation process continued without that task. After the evaluations, 90% of the users made proposals for improvements to the system. The most suggestions were related to the better designing (57%) y improvements with the functions that Google maps use (21%).

4.1 Technical Report

As result of this research, a technical report was prepared with the proposed improvements to guarantee a higher level of usability of the I-Pat system. Table 5 shows the errors found with the respective technical improvement. The improvements were prioritized taking into account the criteria of functionality as priority 1, with the aim that the system is fully functional avoiding the problem of the user being unable to perform their tasks. Priority 2 took into account errors related to clarity and ease of use of the system, i.e. the implementation of user aids, intuitive terms, descriptive icons, help and validation messages, thus avoiding confusion for the user. Finally, taken into account were errors related to interfaces, which are errors that do not stop the operation of the systems, however they must be considered so that the user is comfortable with the use of the system.

Table 5. Technical Improvements

Problems	Improvement
Zoom in maps not working	Check the Google API integration in the application, also enable the zoom functions within the maps.
Patrol does not work	Review the implementation of the k-means algorithm to correctly perform patrolling
Buttons without functionalities	Assign the methods correctly to each button within the application.
The menu does not work well	Verify that the classes and methods assigned to the menu perform the tasks according to the requirements. In addition, it must verify that the corresponding task is enabled.
Messages are not displayed	Verify that the confirmation and validation messages are implemented to facilitate the tasks performed by the user within the application
Non-intuitive functions	Implement messages in each of the functions detailing their main objective
Map without help	The maps must include aids that allow entering the current location, destination and the time it will take for a certain route. In addition it must allow the visualization of alternative routes
Titles on the buttons without sense	The words on the buttons should contain descriptive names, allowing the user to understand what the purpose of that function is. Or, in turn, a link to the user manual must be included within the application.
Non-intuitive terms	Should use terminology that the user can understand. It is recommended to change the labels to more descriptive terms.
There is no help for the user	It is recommended to have a user manual, place tool tip text on the buttons, displaying the contact information to support users.
Icons without description	This problem is related to points 8, 9 and 10. It is also recommended that the buttons, menus and images contain a label with a descriptive name.
Misplaced buttons	The buttons must be located correctly, without blocking access to other functions within the application.
Unsuitable colors	It is recommended to follow some basic principles for handling colors such as theory, models, color schemes that allow, through a study, to determine the colors that help the user feel comfortable with the application
Incorrect menu distribution	It is recommended to prioritize the order of the functions in the menu, you must place the functions with which you work continuously until you reach the lowest level with little-used functions.
Personalization is not allowed	It is recommended to implement a personalization of aspects such as: colors, tones, font types and language with which the user feels more comfortable and can more easily access the application

5 Discussion

As can be seen from the evaluation results, fifty-two errors reported by twenty-one users were found. After the analysis and categorization, these were summarized to fifteen errors in Table 4. The errors that had more incidence were those of comprehension since there were functions of the system that were not correctly implemented, hindering the navigability and completion of the tasks. Secondly, there were the problems related to the interfaces specifically as regards colors, help in the icons and correct distribution of the elements. As can be seen in Table 1, greater errors were obtained using FG. This was because users felt more comfortable expressing their ideas in a group because they

could verify that the other users expressed the same or similar problems. It should be noted that TA is a method that seeks users to express their thoughts aloud. However, it is an activity that users are not used to and it caused some discomfort, in spite of the evaluator, even though he was limited to reminding the users to keep talking.

In addition, there was a questionnaire that had to be completed outside the test site, the questions were aimed at fine-tuning the results obtained in the user test in aspects related to the utility, ease, speed, effectiveness and satisfaction of the system. What users valued most positively was the ease of use of the system, the clarity of the vocabulary and the availability of the application. On the other hand, what the users found most negative was the lack of clarity of the icons, the texts of the menus and the structure of the screens and, finally, the absence of a help page. Regarding the severity of the problems encountered, both with the interface and with comprehension, the highest index is categorized as severity 3, since no problem was identified that would prevent the use of the application from continuing. This, to a certain extent, was influenced by the experience of the users in computer applications.

The literature review suggests the use of TA and its variations for the evaluation of usability [20],[16]. On the other hand, [24]'s research proposes the use of FG as a qualitative research technique widely disseminated in various fields of research in psychology and other scientific disciplines. The present research combines the advantages of both methods to find a greater number of usability problems. As you can see in the results, 5, there were problems that were reported with both methods and others that corresponded only to TA and FG respectively. Due to the nature of each method, the results will be influenced by how the user feels when expressing their thoughts and findings. Some prefer to internalize and analyze their experience before using the system and then express it and discuss it with more users, while others have the ability to instantly communicate what they think while using and testing the system. The purpose of the usability testing studies is to identify system usability problems and provide solutions to solve these problems. In this regard, the users made suggestions for solving these problems and improving system performance. Most of these suggestions were about the better design of some buttons, for example, the close session button, in addition to including aids or descriptive texts in the icons of the different functions. One limitation that was found in the investigation was the lack of users of the National Police, who had provided more objective suggestions on the usability of the system. This was a consequence of the nature of the project since I-Pat is a private government system.

6 Conclusions

Safety is an important necessity in the development of society. For this reason, the National Police of Ecuador sought a solution called I-Pat that, supported by technology, allows patrolling the areas with more conflicts, thus providing a degree of comfort to the population. Being a system of such importance, evaluating usability allows the user to use the system more easily, becoming more familiar with the functions, thus supporting the work of the national police. Usability testing by itself cannot develop a comprehensive list of defects, for this, it is important to use an appropriate mix of methods. In this context, this research allowed a new way of evaluating usability by combining TA

and FG, methods that combined their advantages allowing them to cover a greater number of problems within the system. TA protocols were propagated and far more often used as a method for system usability testing, then they are used to gain insight into end users' thought processes in interaction with a system to support the development of usable systems by identifying system deficiencies. The application of TA and FG in the early stages of the software life cycle will bring great advantages since the software programs, webs or consumer products can be designed with greater precision, not intuitively anticipating the actions of the client, but predicting them quite accurately. Using methods or techniques in a complementary manner as was observed in the research, was an adequate complement, which, when used prior to the debate with the FG, allowed us to obtain an accumulation of qualitative results, which contributed structural elements for the subsequent debate and, at the same time, captured trends in the concepts of the experts. It is noteworthy that FG expose the process, but do not impose specific tools. However, there is a diversity of resources that can be used in each of its phases, which make it versatile and flexible.

References

1. Standards, I.: 9126 standard (2018), <https://www.cse.unsw.edu.au/cs3710/PMmaterials/>
2. ISTQB: International software testing qualifications board, advanced level syllabus usability tester. (2016), <https://www.istqb.org/>
3. Brhel, M., Meth, H., Maedche, A.,Werder, K.: Exploring principles of user-centered agile software development: A literature review. *Information and Software Technology* 61, 163-181 (2015)
4. Pilco, H., Sanchez-Gordon, S., Calle-Jimenez, T., Rybarczyk, Y., Jadán, J., Villarreal, S., Esparza, W., Acosta-Vargas, P., Guevara, C., Nunes, I.L.: Analysis and improvement of the usability of a tele-rehabilitation platform for hip surgery patients. In: *International Conference on Applied Human Factors and Ergonomics*.pp. 197-209. Springer (2018)
5. Maniega-Legarda, D.: Aplicación de criterios de usabilidad en sitios web: consejos y pautas para una correcta interpretación. *Observatorio TIC: REBIUN Red de Bibliotecas Universitarias* (2006)
6. Association, U.E.P.: The usability body of knowledge (2018), <https://www.usabilitybok.org/think-aloud-testing>
7. Guan, Z., Lee, S., Cuddihy, E., Ramey, J.: The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. pp. 1253-1262. ACM (2006)
8. Olmsted-Hawala, E.L., Murphy, E.D., Hawala, S., Ashenfelter, K.T.: Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data dissemination web sites for usability. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. pp. 2381-2390. ACM (2010)
9. McDonald, S., McGarry, K., Willis, L.M.: Thinking-aloud about web navigation: the relationship between think-aloud instructions, task difficulty and performance. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.vol. 57, pp. 2037-2041. SAGE Publications Sage CA: Los Angeles, CA (2013)
10. Hertzum, M., Hansen, K.D., Andersen, H.H.: Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology* 28(2), 165-181 (2009)

11. Barnum, C.M.: Usability testing essentials: ready, set... test! Elsevier (2010)
12. Van Den Haak, M.J., De Jong, M.D., Schellens, P.J.: Constructive interaction: An analysis of verbal interaction in a usability setting. *IEEE transactions on professional communication* 49(4), 311-324 (2006)
13. Makri, S., Blandford, A., Cox, A.L.: This is what im doing and why: Methodological reflections on a naturalistic think-aloud study of interactive information behaviour. *Information Processing & Management* 47(3), 336-348 (2011)
14. Nørgaard, M., Hornbæk, K.: What do usability evaluators do in practice? an explorative study of think-aloud testing. In: *Proceedings of the 6th conference on Designing Interactive systems*. pp. 209-218. ACM (2006)
15. Molich, R., Ede, M.R., Kaasgaard, K., Karyukin, B.: Comparative usability evaluation. *Behaviour & Information Technology* 23(1), 65-74 (2004)
16. Gulliksen, J., Boivie, I., Persson, J., Hektor, A., Herulf, L.: Making a difference: a survey of the usability profession in sweden. In: *Proceedings of the third Nordic-conference on Human-computer interaction*. pp. 207-215. ACM (2004)
17. O Nyumba, T., Wilson, K., Derrick, C.J., Mukherjee, N.: The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution* 9(1), 20-32 (2018)
18. Plummer-D'Amato, P.: Focus group methodology part 2: Considerations for analysis. *International Journal of Therapy and Rehabilitation* 15(3), 123-129 (2008)
19. Onwuegbuzie, A.J., Dickinson, W.B., Leech, N.L., Zoran, A.G.: A qualitative framework for collecting and analyzing data in focus group research. *International journal of qualitative methods* 8(3), 1-21 (2009)
20. Kontio, J., Lehtola, L., Bragge, J.: Using the focus group method in software engineering: obtaining practitioner and user experiences. In: *Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on*. pp. 271-280. IEEE (2004)
21. Sunikka, A.: Usability evaluation of the helsinki school of economics website. Master's thesis (2004)
22. Nielsen, J.: The use and misuse of focus groups. *IEEE software* 14(1), 94-95 (1997)
23. Obeso, M.E.A., Lovelle, J.M.C., Prieto, A.B.M.: Metodología de medición y evaluación de la usabilidad en sitios web educativos. Universidad de Oviedo (2005)
24. Guevara, C., Jadan, J., Zapata, C., Martínez, L., Pozo, J., Manjarres, E.: Model of dynamic routes for intelligent police patrolling. In: *Multidisciplinary Digital Publishing Institute Proceedings*. vol. 2, p. 1214 (2018)
25. Habibi, M.R.M., Khajouei, R., Eslami, S., Jangi, M., Ghalibaf, A.K., Zangouei, S., et al.: Usability testing of bed information management system: A think-aloud method. *Journal of Advanced Pharmaceutical Technology & Research* 9(4), 153 (2018)
26. Harper, B.D., Norman, K.L.: Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5. In: *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference*. pp. 224-228 (1993)
27. Arh, T., Blazic, B.J.: A case study of usability testing-the sumi evaluation approach of the educanext portal. *WSEAS transactions on information science and applications* 5(2), 175-181 (2008)
28. Brooke, J., et al.: Sus-a quick and dirty usability scale. *Usability evaluation in industry* 189(194), 4-7 (1996)
29. Nielsen, J.: Severity ratings for usability problems. *Papers and Essays* 54, 1-2 (1995)

Assessing the Usability of a Mobile Application for Intelligent Police Patrolling

Doris Tutillo-Sánchez^{a*}, Tania Calle-Jimenez^a and Sandra Sanchez-Gordon^a

^aDepartment of Informatics and Computer Science, Escuela Politécnica Nacional, Quito, Ecuador.

{doris.tutillo, tania.calle, sandra.sanchez}@epn.edu.ec

Doris Tutillo-Sánchez is a student on the masters programme in software at the National Polytechnic School of Ecuador and is a professor in the Department of Basic Sciences. She has five years' experience as a quality control analyst in public and private companies, and has an ISTQB certification.

Tania Calle-Jimenez holds a PhD in applications of informatics from the University of Alicante, Spain and a masters in Geomatics from México. She is researcher and professor in the Department of Informatics and Computer Sciences at the National Polytechnic School of Ecuador. Her research interests are eHealth, tele-rehabilitation, human-computer interaction, software quality, accessibility considerations in software requirements engineering and software testing engineering

Sandra Sanchez-Gordon holds a Ph.D. in Applications of Informatics by University of Alicante, Spain and a MSc. in Software Engineering by Drexel University, USA. She is researcher and professor of the Department of Informatics and Computer Sciences of National Polytechnic School of Ecuador. She is Ecuador Representative to Hispanic America Software Testing Qualifications Board (HASTQB). Her research interests are eHealth, tele-rehabilitation, Human-Computer Interaction, MOOCs, software quality, accessibility considerations in software requirements engineering and software testing engineering.

Assessing the Usability of a Mobile Application for Intelligent Police Patrolling

The crime rate has increased in Ecuador. Violations, femicides, homicides and assaults, are all having a negative impact on society. For this reason, the government of Ecuador has developed a mobile application based on Android technology known as I-Pat, which stands for Intelligent Police Patrolling. This research summarises the evaluation of the usability of this application which utilises information about crime reported in three areas over a four-year period. The research includes the application of three techniques: Think Aloud (TA) and Focus Group (FG) techniques are used to evaluate usability, and the NASA Task Load Index (NASA-TLX) is used to measure workload. The findings have led to a number of recommendations being made to improve the application's functionality and the user's experience of it, and so it can be used to support the patrol work of the National Police of Ecuador.

Keywords: Think Aloud; Focus Group; Usability; Patrolling; I-Pat; NASA-TLX; Crime

1. Introduction

The International Organisation for Standardisation (ISO/IEC 25010, 2011) defines the usability of a software product as something that can be learned, understood and used as well as being attractive to the user, when used under specific conditions. However, this concept does not receive the right level of importance when the software is being developed, leading to confusion, errors, delays and the failure of the user to complete certain tasks. Currently, efforts are being made to improve the security, performance, and reliability of software products (Sherman, 2006; Jiang et al., 2016). Unfortunately, too little attention has been paid to addressing software usability problems.

However, this is now changing, resulting in evaluations taking place in the initial phase of the software development process and more active participation from the user in the later phase of the cycle. Usability is now considered a crucial non-functional requirement for economic success in competitive markets (Brhel et al., 2016). There is now the possibility of using specific protocols and instruments to measure the usability level in any type of software (Ivanc and Onita, 2015).

Mobile phones are popular and used regularly in daily life. Due to the emerging deployment of mobile technology, several studies have focused on its usability (Brewster and Murray, 2000; Ryu, 2005). Coursaris and Kim (2006) define four contextual dimensions to consider when evaluating the usability of mobile devices: user, environment, technology and task/activity. Biel, Grill, and Gruhn (2010) applied a hybrid technique to identify usability issues in their research. They said that evaluations of usability should focus on problems using the application and human error. Their study proposed several dimensions, namely usefulness, error, understandability, learnability, satisfaction and intuitiveness. Mobile applications tend to have complex interactive interfaces which result in usability issues such as information overload, screen clutter, lack of task support and limited interaction mechanisms (Wesson, Singh and Van Tonder, 2010).

In this context, this research complements the work of Calle-Jimenez et al. (2019). These authors performed a preliminary evaluation of the usability of the Intelligent System of Police Patrol (I-Pat) developed by the Ministry of the Interior in Ecuador. This application covers one of the fundamental needs of the Ecuadorian community, which is safety supported by technology, including the patrol and surveillance of cities to identify crime and criminal groups. I-Pat is based on a model that generates dynamic routes by applying artificial intelligence through the K-means algorithm to identify critical points

when patrolling Ecuadorian territory (Guevara et al., 2018). In addition, I-Pat incorporates the Google Maps API tool to design routes and verify means of transport. The preliminary usability evaluation of I-Pat measured the system's capacity to be learned, understood, and used by police officers. The authors evaluated the initial phase of the usability of the application based on its usage by 21 students from the Faculty of Systems at the National Polytechnic School using the Think Aloud (TA) and Focus Group (FG) techniques.

In the present study, the evaluation was conducted by computer science professionals and active members of the National Police of Ecuador using TA and FG methods for usability evaluation and NASA-TLX to measure the workload.

This application is currently in the testing phase. Assessments are required to ensure it supports police activities during daily patrols as well as ensuring that the usability level means it can be used without difficulty, otherwise, it cannot be deployed.

The usability was assessed using TA and FG methods to emphasise certain aspects. These include the global structure of the site, content and navigability, without complex terminology which guides the user in an easy to understand manner. In addition, the NASA-TLX method was included to discover any extra dimensions of usability. This method assesses the user's subjective experience of the overall workload and the factors that contribute to it (Hart and Staveland, 1988).

TA is used to evaluate usability and gain an insight into how people work with a product or interface (Guan et al., 2016). It is a method of getting users to verbalise their thought processes as they perform tasks. Often an observer is present to induce the user to be more vocal during this process. It is useful for identifying problems and is relatively simple to configure. In addition, it can provide an indication of the user's attitude, which cannot normally be discerned from a survey or questionnaire (Sánchez, 2015).

FG is a tool that is commonly used in marketing research to determine customers' attitudes about products or services (Caplan, 1990). Participants in an FG session should feel that the process is free-flowing and relatively unstructured, while in reality, the moderator follows a pre-planned script to cover specific issues and sets goals for the type of information to be gathered (Nielsen, 1997). This method consists of a planned and designed discussion about the interfaces. In this research, the discussion was guided and facilitated by a moderator who followed a pre-defined structure so that the discussion was focused. A brainstorming activity was carried out so that the participants could exchange opinions.

NASA-TLX is a subjective workload assessment technique that relies on a multidimensional construct to derive the overall workload. The score is based on a weighted average of ratings from six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration level. Three of the subscales relate to the demands imposed on the user (mental, physical, and temporal demand), whereas the others focus on the interaction of the user with the task (performance, effort, and frustration level) (Cao, Chintamani, Pandya and Ellis, 2009). The NASA-TLX method has been used in several fields (Hart, 2006) including studies involving evaluations of visual and auditory displays, vocal and manual input devices, and virtual and augmented vision. The researchers explored the relationship between NASA-TLX ratings and other performance factors, such as fatigue, stress, trust, experience and situational awareness. Due to its wide applicability, it has also been used in combination with other methods for the evaluation of usability (Park, Harada, and Igarashi, 2006).

The mixture of TA and FG usability methods provide a feasible approach for identifying important usability aspects and for encouraging end-user participation with the application. The TA technique is useful for identifying the detailed requirements

relating to the user interface. The FG interviews, on the other hand, are important for incorporating the contextual issues. In addition, the authors applied the NASA-TLX method. This allows the workload to be measured and is sensitive to the complexity of the tasks evaluated by the final users.

Finally, a report has been produced that presents the usability errors and makes technical recommendations. The report was delivered to the National Police of Ecuador for the subsequent improvement of the I-Pat system.

2. Materials and Methods

This study uses the TA and FG usability assessment methods and NASA-TLX to measure the workload of a police patrol application (I-Pat), developed by the National Police of Ecuador. I-Pat, the participants in the assessment and the assessment process are described in detail in this section.

2.1. Mobile Application I–Pat

I-Pat is an acronym for Intelligent Police Patrolling. It is a mobile application which has been developed to deal with one of the most fundamental needs of the Ecuadorian community - safety. The application provides support to the police when patrolling and surveilling Ecuadorian cities and helps them identify crime and criminal groups. It uses information about reported crime in three areas over a four-year period (Guevara et al., 2018) and is supported by a model that generates dynamic routes by applying artificial intelligence using a K-means algorithm to identify critical points when patrolling the Ecuadorian territory. I-Pat includes historical modules with up-to-date information about various thefts including those from people, houses and banks, and the theft of vehicles, accessories and on the roads. It also has a complementary section with functions that allow the police to request reinforcement

patrols, predictive maps, agents' schedules and to report civil unrest. Figure 1 shows the I-Pat interface which displays the information about police officers, the summary of the completed patrols and a map of the circuit covered by the police officer.

2.2. Participants

The assessment results in this research were obtained from participants in two test groups. The first group was the experimental group. This group of 18 people was made up of computer science professionals and students in their final year of the computer science course. These participants had the ability to design and analyse computer systems as well as being able to understand the principles and concepts of programming languages. They had no knowledge or experience of the I-Pat mobile application. The second group was the control group made up of five participants who were active members of the National Police of Ecuador. According to Nielsen (2012), the best results are obtained from testing no more than five users and by running as many small tests as possible, because after the fifth user, they say that time is wasted observing the same findings repeatedly and nothing new is learned. The participants had knowledge of mobile devices and would be the main users of I-Pat. Before starting the evaluations, all participants were familiarised with the objectives of the study, and their informed consent was obtained. The evaluation was carried out in two sessions. The TA method was applied in the first session; the evaluator acted as the facilitator together with each user. The facilitators did not intervene in the evaluation process, but were there to help if users halted the usability testing. The facilitators reminded the users to express their thoughts out loud (Van Someren et al.,1994). In the second session, the FG method was used and the facilitator took on the role of moderator, directing the groups to obtain the greatest number of ideas about the usability of I-Pat.

2.3. *Assessment Methods*

2.3.1. *Think Aloud*

The TA method has its roots in psychological research. It was developed from the older introspection method. Introspection is based on the idea that one can observe events that take place in consciousness, as one can observe events in the outside world (Van Someren and Sandberg, 1994). The TA method is a tool that supports usability testing and is classified among the main methods used by usability professionals for prototyping, interviews, field studies and in other scenarios (Gulliksen, Boivie, Persson, Hektor, and Herulf, 2004).

It is impossible to directly observe what a user is thinking, so TA is frequently used for this reason. By using TA, those areas where a user struggles and the reasons for their difficulties are verbally articulated. The usability evaluator uses this information along with other measures to identify the problem areas of the application being assessed and to devise suggestions for improvement. The participants are encouraged to think out loud while working on a task.

2.3.2. *Focus Group*

The FG research method emerged in social research in the 1950's when scholars expanded the open ended interview format into group discussions (Templeton, 1994). Using FGs is a method of collecting qualitative data, and essentially involves engaging a small number of people in an informal group discussion, focused around a specific topic or set of issues (Wilkinson and Silverman, 2004).

FG are frequently used as a qualitative approach to gain an in-depth understanding of social issues. The method aims to obtain data from a purposely selected group of

individuals rather than from a statistically representative sample of a broader population. It includes carefully planned discussions, designed to obtain the perceptions of the group members about a defined area of interest. There are typically between three and 12 participants and the discussion is guided and facilitated by a moderator, who follows a pre-defined structure so that the discussion remains focused. The members are selected based on their individual characteristics related to the session topic. The group setting enables the participants to build on the responses and ideas of others, which increases the richness of the information (Kontio, Lehtola, and Bragge, 2004).

2.3.3. *NASA-TLX*

NASA-TLX is a multidimensional assessment procedure, in which a global workload score is obtained based on the weighted average of six variables or sub scales, which are: mental demands, physical demands, temporal demands, effort, frustration and performance. The idea of NASA-TLX is to allow the user time to access the workload situation once testing is complete. The subscales provide detailed information, not just one-answer responses. NASA-TLX does not test the user while they are in the process of completing the task. The user relies on what they remember and provides an opinion based on memory (Tracy and Albers, 2006).

The assessment has two phases. In the first phase, the users assign weight to every dimension before the execution of the task. In the second phase, the users assign a score to every one of the six dimensions immediately after the execution of the task. (Pilco et al., 2019). The main goal is to predict the workload generated by the web interfaces, thus adding one more dimension of usability to the mobile application, and guaranteeing greater user safety. The NASA-TLX method provides an overall score based on a weighted average of ratings of six subscales: mental, physical and temporal

demand, effort, performance and frustration level. A description of each of these can be found in Table 1.

2.4. Tools and Environment

The assessment of the experimental group was carried out at the National Polytechnic School of Ecuador. For the control group, the assessment was carried out in the Mariscal Sucre Regiment facility at the National Police of Ecuador. The assessment process was performed using DU recorder software to capture user interaction with I-Pat. The microphone and video camera were used to record the user's voice and capture their gestures. The assessment process included two tasks: 1) Historical navigation; where the information about crime committed in the city is shown, 2) Complements module; this displays the information related to patrols, police personnel schedules and the concentration of the crime.

3. Results

3.1. General Results

Figure 2 shows the process followed to evaluate the usability of I-Pat using the TA and FG methods and the NASA-TLX method to measure the workload. Once the testers started the application, the users were introduced to and familiarised with the system's functions.

The first session used the TA method to collect information about the users' cognitive interaction with the system (Habibi et al., 2018). The users were asked to express what they saw, thought and felt with respect to the use of the I-Pat interfaces. After the introduction, the TA assessment lasted 15 minutes. During this session the facilitator reminded the users to express their thoughts out loud and audio recorded them for later analysis.

The FG method was used in the second session. For the experimental group, two sessions were organised based on the number of participants. However, for the control group only one session was executed. This is in accordance with the recommendation to determine the focus group size, the number of questions asked, the time allotted for each question, the format of the focus group session and the duration of the session (Tang and Davis, 1995). The moderator brainstormed with the participants and during the free exchange of ideas between the testers obtained any original ideas arising from the evaluation of the usability of I-Pat. The moderator explained the rules of the session, including the need to express ideas and avoid criticism. The moderator also explained that all thoughts were important, even if they did not appear to make sense because they might serve as inspiration for new developments. The testers expressed their ideas to the rest of the group.

The experimental group participants included 60 (60%) females and 40 (40%) males whose ages ranged from 19-27 years (25%), 28-34 years (35%), and over 32 years (40%). In the group using the TA method 21 errors were found. However, some errors were reported several times by different users. Most of the errors found related to incorrect or unintuitive terms, the lack of textual aids and user manual. In addition, some functionality errors were reported. In the group using the FG method, 34 errors were reported, including functionality, a lack of help and interface errors. To summarise, 52 errors were reported by 18 users but after analysis and categorisation, they were summarised into 18 total errors. All errors found were classified according to the Haak method (Van Den Haak et al., 2006). This method establishes four main categories of problem: layout, terminology, data entry and comprehensiveness.

In addition, the Nielsen method (Nielsen, 2017) was used to identify the severity of the problems found, assigning the values of zero for the absence of usability errors,

one for cosmetic problems, two for minor problems, three for major problems and four for catastrophes related to usability.

The classification and severity of the errors encountered by users during the assessment process in the experimental group is shown in Table 2. According to Table 2, the problem of comprehensiveness was the most frequent with seven (39%) cases. Around 71% of the problems in this category had a severity of three. In this category, the main problems reported related to functions such as the patrol not working (12), non-descriptive icons (18), and functions whose purpose was unknown (6). Five (28%) data entry problems were also identified by users. In this category, three problems were rated as severe; the zoom malfunctioned on the maps, messages were not displayed, and the menu did not work very well. These were the most recurrent problems.

After applying the FG method, the main findings relate to the users' experience operating I-Pat; 15 (83%) users had no problems using the application and three (17%) experienced between one and three interruptions without stopping the evaluation process.

Figure 3 shows the results obtained from the application of NASA-TLX. It shows that the global weighted average for the two tasks did not exceed 50%, therefore performing these tasks has a low to medium psychosocial risk factor relating to load and mental fatigue.

From the data obtained after analysis, it is observed that for both tasks the mental, physical and temporal demand is very low, as it is not necessary to memorise any of the procedures or make any physical effort. In addition, the rhythm of execution of the task can be paused without pressure (see Figure 4). The results show that with some effort the participants met the objectives set for the evaluation in task one.

However, for task two they had to increase their effort, and the objectives set were not achieved. This is consistent with the findings of the tests performed using the TA method where it is evident that users experienced more usability errors related to problems of comprehension.

As the five participants in the control group were police officers who have worked in the northern part of the city and in the Quito valleys they already had some knowledge of the mode of operation for patrolling and of the criminal acts. The participants included four females (80%) and one male (20%) with ages ranging from between 28 and 35 years. Using the TA and FG methods resulted in 30 errors; however, some errors were reported several times by different users.

Most errors found in the TA assessment related to incorrect or unintuitive terms, the lack of textual aids and the interface. In the FG assessment, most errors related to functionality, specifically regarding the information about delinquent acts. To summarise, 30 errors were reported by five users, but after analysis and categorisation, they were summarised into twelve errors. Table 3 shows the classification and severity of the errors that users encountered during the assessment process.

According to Table 3, problems of comprehensiveness were the most frequent with seven (58%) cases. Approximately 71% of the problems in this category had a severity of three. In this category, the main problems reported were about functions not working such as the patrolling, or about non-descriptive icons, functions with an unknown purpose and information not being updated.

Three (25%) data entry problems were identified by users and rated as severity three, including the zoom in the maps malfunctioning, messages not being displayed, and the menu not working well. These were the most recurrent problems.

After applying the FG method, the main findings related to users' experience operating I-Pat with the main error relating to information in the system. Each module should refresh the information about criminals and the areas with the most criminal activity. In addition, the patrolling option must show the reinforcements closest to the emergency.

Figure 5 shows the results obtained from the application of the NASA-TLX method in the control group. Similar to the findings for the experimental users, the global weighted average for the two tasks does not exceed 50%, therefore performance of these tasks has a psychosocial risk factor of low-medium equivalent to the load and mental fatigue.

From the data obtained after analysis it can be observed that for both tasks the physical and temporal level of frustration is very low, as it is not necessary to memorise any procedure, make any physical effort, and the rhythm of execution of the task can be paused without pressure (see Figure 6). The results show that the participants met the objectives set in tasks one and two which is likely due to the users having the knowledge and experience of the logic of the application.

3.2. Technical Report

As a result of this research, a technical report has been prepared detailing the proposed improvements that would guarantee a higher level of usability of I-Pat. Table 4 lists the errors found and their respective technical improvements. The improvements have been prioritised based on the criteria of functionality as priority one. The aim is to ensure I-Pat is fully functional and problem free so that the user can perform their tasks. Priority two relates to errors about clarity and ease of use such as the implementation of user aids, intuitive terms, descriptive icons, help and validation messages, which help to avoid user confusion. Finally, some errors relate to the

interfaces but do not prevent the application from being operated; however, the user must be comfortable using the system so these errors must be considered.

Figure 7 shows a proposal for the application's patrolling interface. It suggests three modifications: allow a zoom facility for the map, reinitialise the map according to the user's position, and relocate the close session button at the bottom.

Additionally, it is recommended that training on the functionalities of I-Pat is provided to users who have to interact with the application, in order to reduce the levels of frustration and increase the levels of performance as obtained in the evaluation using NASA-TLX.

4. Discussion

Today mobile technology is used almost everywhere and its usage is increasing. It is used in government administration to deliver public services to citizens, and by civil servants and public and private firms. The government of Ecuador has prioritised safety as the primary need of its citizens, and focused its efforts on supporting the work of the National Police, by using new technology to reinforce the citizen safety service. The mobile application I-Pat has been developed as a tool to support the tasks of patrolling police officers and to manage extensive information about criminal circuits, crime and data related to criminals. The evaluation of the usability of the mobile application I-Pat aims to minimise the impact of any potential usability problems and thereby allow more efficient use by the end users. For this reason, it was important to carry out a usability assessment study to motivate police officers to use the mobile application safely, effectively, efficiently, and satisfactorily.

This study uses two different usability evaluation approaches as well as the NASA-TLX method for workload measurement. The two different types of usability

analyses provide a multi-faceted understanding of what users expect from application tasks and how they expect to interact with the application. The literature review suggests the use of the TA method and its variations to evaluate usability (Gulliksen et al., 2004; Olmsted-Hawala et al., 2010). On the other hand, Kontio et al. (2004) proposed the use of the FG method as a qualitative research technique which is widely used in various fields of research including psychology and other scientific disciplines. This research benefits from the advantages of both methods to determine a greater number of usability problems. As the results show, problems were reported using both methods with some corresponding only with TAs and some with FGs. Due to the nature of each method, the results are influenced by how the users feel when expressing their thoughts and findings. Some users prefer to internalise and analyse their experience before using the mobile application and then express themselves and discuss this with other users. Whereas others can instantly communicate what they are thinking while using and testing the mobile application.

In addition, the subjective ratings of mental workload were measured using NASA-TLX. In the research by Harrison and Duce (2013), the authors mention the importance of cognitive load as an attribute of usability. In that study it was found that almost 23% of studies measured the cognitive load of the application being evaluated. The NASA-TLX score ranged from zero (no effort) to 100 (maximum effort) (Hart and Staveland, 1988). With the application of NASA-TLX it was possible to demonstrate that workload does not exceed 50% when using the application. However, to achieve the highest performance of the application it is recommended to carry out prior training before using it, so that users can explore the options without difficulty and without getting frustrated as was the case with the experimental users.

The application of TA, FG and NASA-TLX methods provided a team of developers from the National Police of Ecuador with the information they needed to optimise the implementation of the application and predict the cognitive workload. In this regard, we made suggestions for solving the usability problems and improving the performance of the application. Most of these suggestions relate to fixing the functional problems, re-designing some of the buttons, including some aids or descriptive text for the icons for different functions and conducting training for those users who work with the application every day.

Due to restricted access to the application, it was not feasible to carry out another usability evaluation after the implementation of the technical recommendations. However, because of the importance of the application for improving the patrolling service and the commitment of the institution to providing its employees with an application to support their daily tasks, the functional problems are going to be solved and lectures are going to be provided about the use of the application.

5. Conclusion and Future Work

Society values safety. For this reason, the National Police of Ecuador sought a solution known as I-Pat that utilises technology to support patrolling officers in areas of conflict, thus providing a degree of comfort to the population. Being such an important system, a usability assessment has helped the users to learn about and operate the system more effectively and to familiarise themselves with the functions, thus supporting the work of the national police. Usability testing by itself cannot produce a comprehensive list of defects, for this, it is important to use an appropriate mix of methods. In this context, this research allowed a new way of evaluating usability by combining TA and FG methods to assess usability and NASA-TLX to

assess the cognitive workload. Combining these methods has allowed a greater number of problems with the application to be discovered.

This study finds that many of the problems identified during the evaluation process relate to comprehensiveness, which is due to poor system design and a lack of consideration of users' needs. Therefore, it is recommended that users be educated about the application's functions before using it. Carrying out a usability assessment in the early stages of the software life cycle has many advantages since software programs, webs or consumer products can be designed with greater precision. It may not intuitively anticipate all the actions of the client, but could predict them fairly accurately.

For future work, a study could be conducted based on the new version of the application, which has incorporated the improvements proposed in the technical report. Usability experts could be included to provide a focus for the study

6. References

1. Biel, B., Grill, T., & Gruhn, V. (2010). Exploring the benefits of the combination of a software architecture analysis and a usability evaluation of a mobile application. *Journal of Systems and Software*, 83(11), 2031-2044.
2. Brewster, S., & Murray, R. (2000). Presenting dynamic information on mobile computers. *Personal and Ubiquitous Computing*, 4(4), 209-212.
3. Brhel, M., Meth, H., Maedche, A., & Werder, K. (2015). Exploring principles of user-centered agile software development: A literature review. *Information and software technology*, 61, 163-181.
4. Calle-Jimenez, T., Tutillo, D., Sanchez-Gordon, S., Jadán-Guerrero, J., Guevara, C., Lara-Alvarez, P., Acosta-Vargas, P., Salvador-Ulluari, L., Nunez, I., (2019). Improving Usability with Think Aloud and Focus Group Methods. Case Study: An Intelligent Police Patrolling System (I-Pat).
5. Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior research methods*, 41(1), 113-117.
6. Caplan, S. (1990). Using focus group methodology for ergonomic design. *Ergonomics*, 33(5), 527-533.
7. Coursaris, C., & Kim, D. (2006). A qualitative review of empirical mobile usability studies. *AMCIS 2006 Proceedings*, 352.
8. Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006, April). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In

- Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 1253-1262). ACM.
9. Guevara, C., Jadán, J., Zapata, C., Martínez, L., Pozo, J., & Manjarres, E. (2018). Model of Dynamic Routes for Intelligent Police Patrolling. In *Multidisciplinary Digital Publishing Institute Proceedings* (Vol. 2, No. 19, p. 1214). Proceedings of UCAMI 2018, <https://doi.org/10.3390/proceedings2191214>.
 10. Gulliksen, J., Boivie, I., Persson, J., Hektor, A., & Herulf, L. (2004, October). Making a difference: a survey of the usability profession in Sweden. In *Proceedings of the third Nordic conference on Human-computer interaction* (pp. 207-215). ACM.
 11. Habibi, M. R. M., Khajouei, R., Eslami, S., Jangi, M., Ghalibaf, A. K., & Zangouei, S. (2018). Usability testing of bed information management system: A think-aloud method. *Journal of advanced pharmaceutical technology & research*, 9(4), 153.
 12. Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science*, 1(1), 1.
 13. Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.
 14. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.
 15. International Organization for Standardization ISO 25010, Systems and software engineering. Available online: <https://www.iso.org/standard/35733.html>. (2011)
 16. Ivanc, D., Onita, M. (2015) Usability evaluation of a LMS mobile web interface. In *Proceedings of the International Conference on Information and Software Technologies*, 348–361. https://doi.org/10.1007/978-3-642-33308-8_29
 17. Jiang, Y., Wu, D., Liu, P. (2016) JRed: Program Customization and Bloatware Mitigation Based on Static Analysis. In *Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference* (1), 12–21. <https://doi.org/10.1109/compsac.2016.146>
 18. Kontio, J., Lehtola, L., & Bragge, J. (2004, August). Using the focus group method in software engineering: obtaining practitioner and user experiences. In *Proceedings. 2004 International Symposium on Empirical Software Engineering, 2004. ISESE'04.* (pp. 271-280). IEEE.
 19. Maniega-Legarda, D.: Aplicación de criterios de usabilidad en sitios web: consejos y pautas para una correcta interpretación. [Application of usability criteria in websites: tips and guidelines for correct interpretation] Observatorio TIC: REBIUN Red de Bibliotecas Universitarias (2006)
 20. Nielsen, J. (2012). Why you only need to test with 5 users, 2000. Jakob Nielsen's Alertbox. Available on line from www.useit.com/alertbox/20000319.html.
 21. Nielsen, J. (2017). Severity Ratings for Usability Problems. Nielsen Norman Group.
 22. Nielsen, J. (1997). The use and misuse of focus groups. *IEEE software*, 14(1), 94-95.
 23. Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010, April). Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2381-2390). ACM.

24. Park, S., Harada, A., & Igarashi, H. (2006, April). Influences of personal preference on product usability. In CHI'06 Extended Abstracts on Human Factors in Computing Systems (pp. 87-92). ACM.
25. Pilco, H., Sanchez-Gordon, S., Calle-Jimenez, T., Pérez-Medina, J. L., Rybarczyk, Y., Jadán-Guerrero, J., ... & Nunes, I. L. (2019). An Agile Approach to Improve the Usability of a Physical Telerehabilitation Platform. *Applied Sciences*, 9(3), 480
26. Ryu, Y. S. (2005). Development of usability questionnaires for electronic mobile products and decision making methods (Doctoral dissertation, Virginia Tech).
27. Sánchez, W. O. (2015). La usabilidad en Ingeniería de Software: definición y características [Usability in software engineering: definition and characteristics]
28. Sherman, P. (2006) How Organizations Improve By Making Easier-To-Use Software and Web Sites. *Usability Success Stories*. NY, USA: Gower Publishing <https://doi.org/10.1080/00140130802048476>
29. Tang, K. C., & Davis, A. (1995). Critical factors in the determination of focus group size. *Family practice*, 12(4), 474-475.
30. Templeton, J. F. (1994). *The focus group: a strategic guide to organizing, conducting and analyzing the focus group interview*. Probus Publishing Company.
31. Tracy, J. P., & Albers, M. J. (2006, May). Measuring cognitive load to test the usability of web sites. In *Annual Conference-society for technical communication* (Vol. 53, p. 256).
32. Van Den Haak, M. J., De Jong, M. D., & Schellens, P. J. (2006). Constructive interaction: An analysis of verbal interaction in a usability setting. *IEEE transactions on professional communication*, 49(4), 311-324.
33. Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: a practical approach to modelling cognitive*. Academic Press, London.
34. Wesson, J. L., Singh, A., & Van Tonder, B. (2010, September). Can adaptive interfaces improve the usability of mobile applications? In *IFIP Human-Computer Interaction Symposium* (pp. 187-198). Springer, Berlin, Heidelberg.
35. Wilkinson, S., & Silverman, D. (2004). 10 Focus Group Research. *Qualitative research: Theory, method and practice*, 177-199.

Figure 1: The Patrolling Interface of the Mobile Application I-Pat 21

Figure 2: The Usability Assessment Process in I-Pat..... 22

Figure 3: NASA – TLX Global weighted average- Experimental User Group 23

Figure 4: NASA-TLX Weighted Score in the Experimental User Group..... 23

Figure 5: NASA-TLX Global Weighted Average in the Control Group of Users..... 24

Figure 6: NASA-TLX Weighted Score in the Control Group of Users 24

Figure 7: Interface before and after improvement..... 25



Figure 1: The Patrolling Interface of the Mobile Application I-Pat

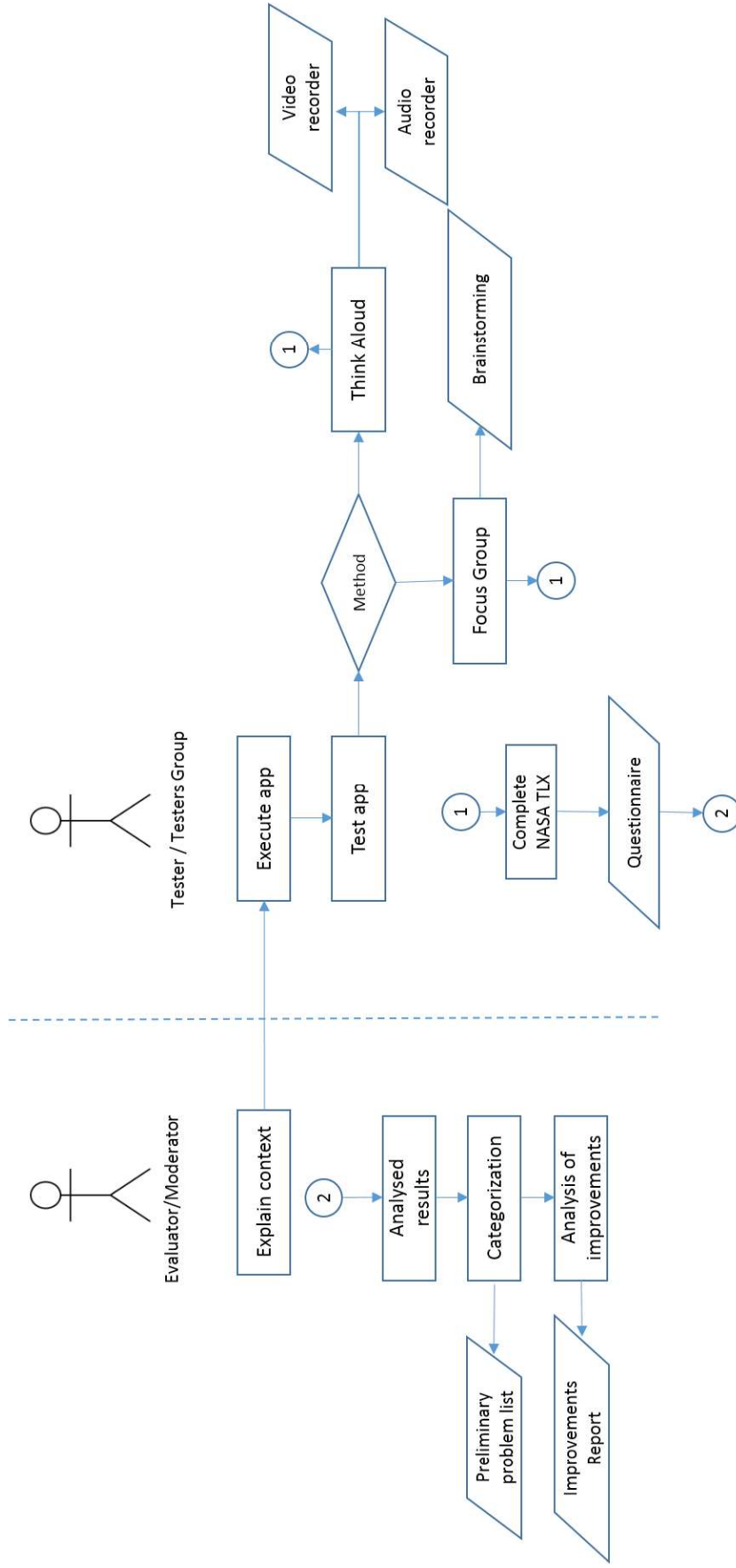


Figure 2: The Usability Assessment Process in I-Pat

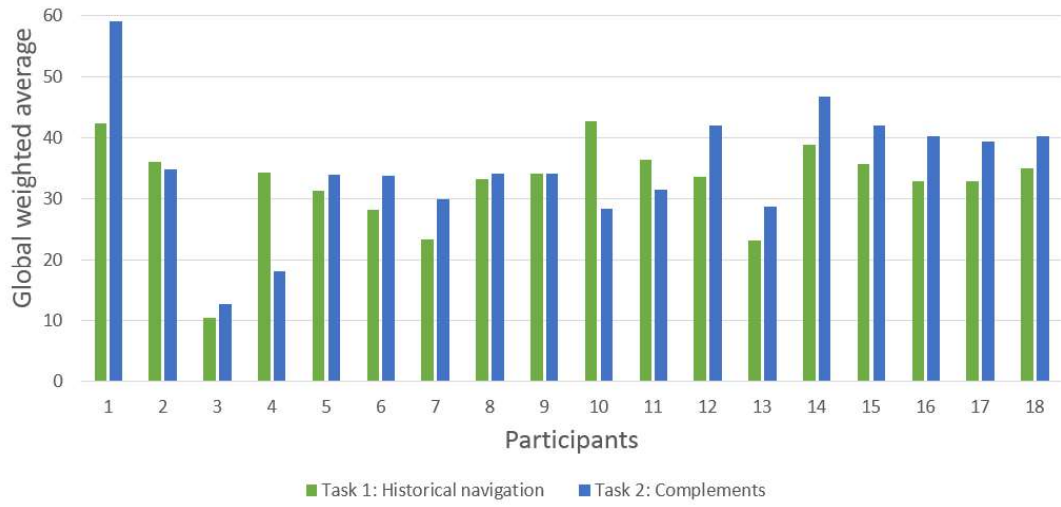


Figure 3: NASA – TLX Global weighted average- Experimental User Group

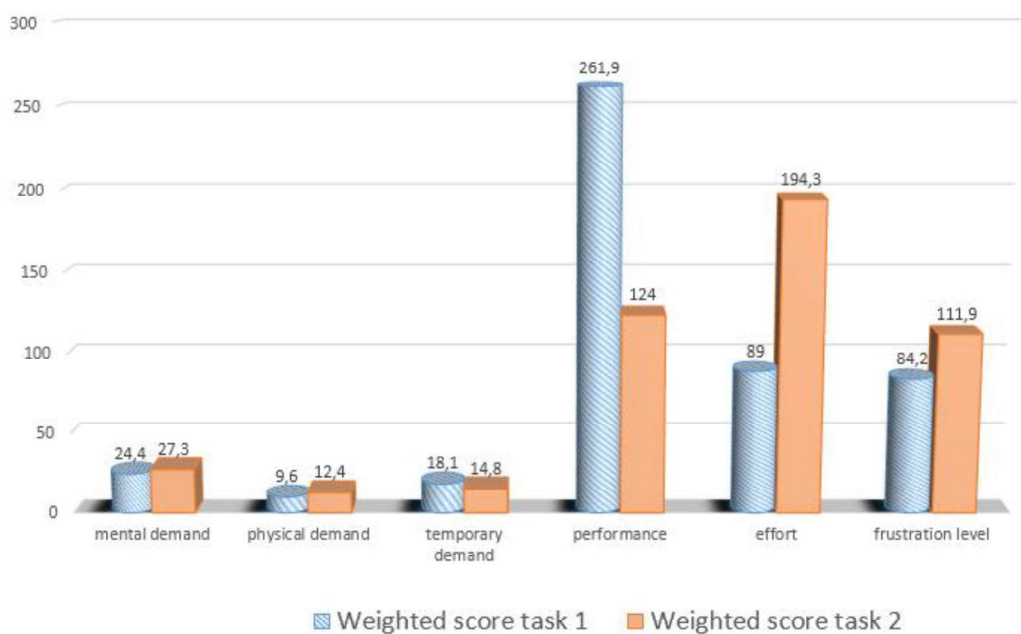


Figure 4: NASA-TLX Weighted Score in the Experimental User Group

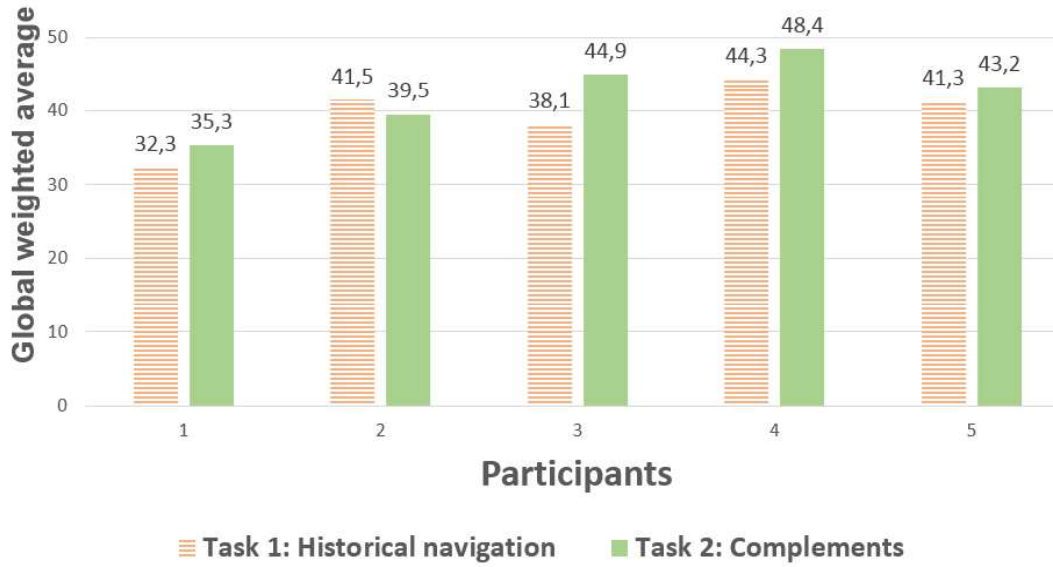


Figure 5: NASA-TLX Global Weighted Average in the Control Group of Users

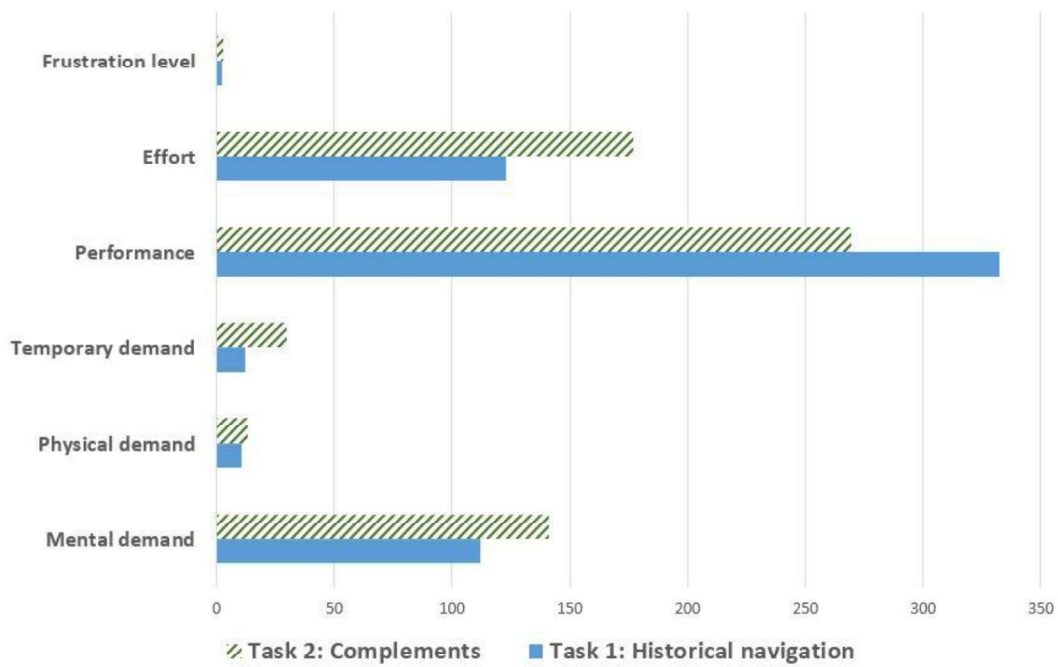


Figure 6: NASA-TLX Weighted Score in the Control Group of Users

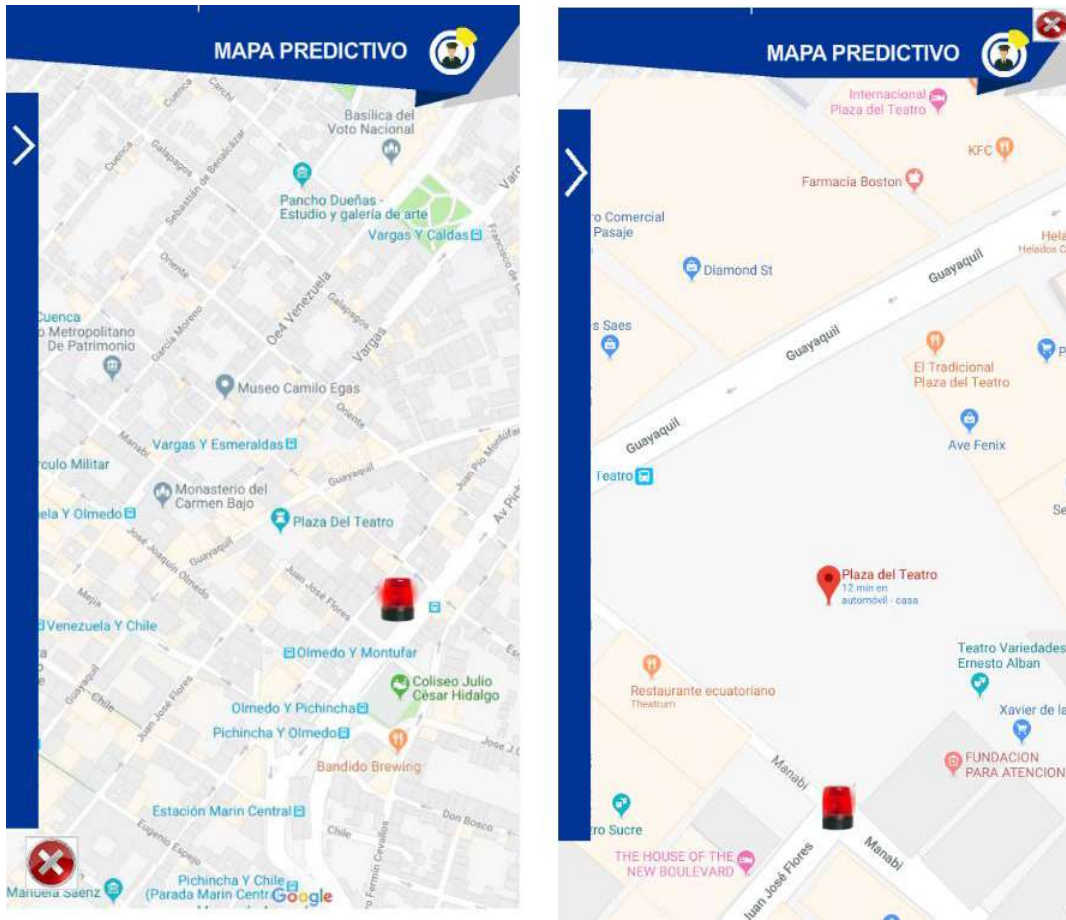


Figure 7: Interface before and after improvement

Table 1: Rating scale definitions of NASA-TLX (Hart & Staveland,1988).	27
Table 2: Experimental Group's Usability problems using TA and FG	28
Table 3: Control Group's Usability problems using TA and FG	29
Table 4: Technical Improvements	31

Title	Endpoints	Descriptions
Mental Demand	Low / High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.) was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low / High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, etc.) was the task easy or demanding, slow or brisk, slack or strenuous?
Temporal Demand	Low / High	How much time pressure did you feel due to the rate or pace at which the task or task element occurred? Was the pace slow and leisurely or rapid and frantic?
Effort	Low / High	How hard did you have to work (mentally and physical) to accomplish your level of performance
Performance	Good /Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
Frustration level	Low / High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Table 1: Rating Scale Definitions of NASA-TLX. Source: (Hart and Staveland, 1988).

Item	Problems	Category	Severity	Method
1	Zoom in maps not working	Data entry problem	3	TA/FG
2	Incorrect menu distribution	Layout problems	1	TA
3	Icons without description	Comprehensiveness problems	2	FG
4	Buttons without functions	Comprehensiveness problems	3	TA/FG
5	Titles on the buttons made no sense	Comprehensiveness problems	3	TA
6	Map without help	Comprehensiveness problems	3	TA
7	Misplaced buttons	Layout problems	1	TA/FG
8	Unsuitable colours	Layout problems	1	FG
9	The menu did not work well	Data entry problem	3	FG
10	Personalization was not allowed	Layout problems	1	FG
11	Messages were not displayed	Data entry problem	3	FG
12	Non-intuitive functions	Comprehensiveness problems	3	FG
13	There was no help for the user	Comprehensiveness problems	2	TA/FG
14	Patrolling did not work	Comprehensiveness problems	3	FG
15	Non-intuitive terms	Terminology problems	2	TA/FG
16	Menus overlapping	Data entry problem	2	TA/FG
17	Maps without legends	Terminology problems	1	TA
18	Operating system restriction	Data entry problem	1	TA/FG

Table 2: Experimental Group's Usability problems using TA and FG methods

Item	Problems	Category	Severity	Method
1	Non-intuitive terms	Terminology problems	2	TA
2	Buttons without information/labels	Comprehensiveness problems	2	TA
3	Zoom in maps not working	Data entry problem	3	TA/FG
4	Buttons without functions	Comprehensiveness problems	3	TA
5	Patrolling did not work	Comprehensiveness problems	3	TA/FG
6	Messages were not displayed	Data entry problem	2	FG
7	Buttons without help	Comprehensiveness problems	2	FG
8	The menu did not work well	Data entry problem	2	TA/FG
9	Maps without legends	Terminology problems	2	TA
10	Delinquent's information is not updated	Comprehensiveness problems	3	FG
11	Sector's information is not updated	Comprehensiveness problems	3	FG
12	Victims' information is not updated	Comprehensiveness problems	3	FG

Table 3: Control Group's Usability problems using TA and FG methods

	Problems	Improvement
1	Zoom in maps not working	Check the Google API integration in the application, also enable the zoom functions within the maps
2	Delinquent's information is not updated	Verify that the information is updated according to the indicated functionality.
3	Sector's information is not updated	Verify that the information is updated according to the indicated functionality.
4	Victims' information is not updated	Verify that the information is updated according to the indicated functionality.
5	Patrolling does not work	Review the implementation of the k-means algorithm to correctly perform patrolling
6	Buttons without functionalities	Assign the methods correctly to each button within the application.
7	The menu does not work well	Verify that the classes and methods assigned to the menu perform the tasks according to the requirements. In addition, it must verify that the corresponding task is enabled.
8	Messages are not displayed	Verify that the confirmation and validation messages are implemented to facilitate the tasks performed by the user within the application
9	Non-intuitive functions	Implement messages in each of the functions detailing their main objective
10	Map without help	The maps must include aids that allow entering the current location, destination and the time it will take for a certain route. In addition it must allow the visualization of alternative routes
11	Titles on the buttons without sense	The words on the buttons should contain descriptive names, allowing the user to understand what the purpose of that function is. Or, in turn, a link to the user manual must be included within the application.
12	Non-intuitive terms	Should use terminology that the user can understand. It is recommended to change the labels to more descriptive terms.

13	There is no help for the user	It is recommended to have a user manual, place tool tip text on the buttons, displaying the contact information to support users.
14	Icons without description	This problem is related to points 9, 11 and 12. It is also recommended that the buttons, menus and images contain a label with a descriptive name.
15	Misplaced buttons	The buttons must be located correctly, without blocking access to other functions within the application.
16	Unsuitable colours	It is recommended to follow some basic principles for handling colours such as theory, models, colour schemes that allow, through a study, to determine the colours that help the user feel comfortable with the application
17	Incorrect menu distribution	It is recommended to prioritize the order of the functions in the menu, you must place the functions with which you work continuously until you reach the lowest level with little-used functions.
18	Personalization is not allowed	It is recommended to implement a personalization of aspects such as: colours, tones, font types and language with which the user feels more comfortable and can more easily access the application

Table 4: Technical Improvements

Evaluacion de Usabilidad del Sistema I-PAT

Las preguntas descritas en el presente formulario buscan medir el grado de satisfacción que usted como participante del experimento tuvo durante la utilización del sistema. La información aquí suministrada es totalmente confidencial y solo será utilizada para fines académicos relacionados con esta investigación. El cuestionario cuenta con 17 preguntas. Por favor responda cada una de ellas seleccionando el valor en la escala de 0 y 9 que para usted es el más apropiado para cada pregunta.

1. Dirección de correo electrónico *

2. Ingrese su rango de edad

Selecciona todos los que correspondan.

- 18- 22
 23-27
 28-32
 Mayor a 32

3. Como considera sus conocimientos en Ciencias de la Computación

Marca solo un óvalo.

	1	2	3	4	5	
Bajos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Altos

4. Especifique su género

Marca solo un óvalo.

- Femenino
 Masculino

APRENDIZAJE

5. 1. Aprendiendo a manejar el sistema

Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	
Difícil	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fácil

6. 2.Explorando nuevas características por prueba y error*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Difícil	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fácil

7. 3.Recordando nombres y uso de comandos*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Difícil	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Fácil

8. 4. Las tareas pueden ser realizadas de una manera sencilla*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

9. 5. La herramienta fue fácil de usar*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

10. 6. Pienso que la mayoría de la gente aprenderá a usar la herramienta muy rápidamente*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

11. 7. En un inicio aprender a utilizar esta herramienta es difícil*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

Capacidades del Sistema

12. 1. Velocidad del sistema*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Demasiado lento	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muy rápido

13. 2. Confiabilidad del sistema*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Poco Fiable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Confiable

14. 3. El software en algún momento se detuvo inesperadamente*Marca solo un óvalo.*

	1	2	3	4	5	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

Usabilidad e Interfaz de Usuario**15. 1. Uso de colores***Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Deficiente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Excelente

16. 2. Retroalimentación del sistema*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	10	
Deficiente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Excelente

17. 3. Trabajar con este software es mentalmente estimulante*Marca solo un óvalo.*

	1	2	3	4	5	6	7	8	9	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

18. 4. La forma en que el sistema presenta la información es clara y comprensible

Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

19. 5. Puedo comprender y manejar la información proporcionada por el sistema

Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	
Totalmente en desacuerdo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totalmente de acuerdo

Comentarios

20. Permítanos conocer sus impresiones que le ha dejado el uso de la herramienta

Recibir una copia de mis respuestas

Con la tecnología de



NASA-TLX Sistema I-Pat

Task Load Index Índice de carga de trabajo, es un procedimiento de valoración multidimensional, el cual mide la carga de trabajo basado en la media ponderada de seis variables.

*Obligatorio

Datos Demográficos

1. Ingrese su edad

2. Ingrese su género

Marca solo un óvalo.

Femenino

Masculino

3. Ingrese el nivel de experiencia con dispositivos móviles

Marca solo un óvalo.

	1	2	3	4	5	
Bajo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

Ponderación por actividad

Asigne un peso a cada dimensión:

4. Actividad 1: Navegación Históricas *

Marca solo un óvalo por fila.

	1	2	3	4	5
Exigencia Mental	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exigencia Física	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exigencia Temporal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Esfuerzo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rendimiento	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nivel de Frustración	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. Actividad 2: Navegación Complementos **Marca solo un óvalo por fila.*

	1	2	3	4	5
Exigencia Mental	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exigencia Física	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exigencia Temporal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Esfuerzo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rendimiento	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nivel de Frustración	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Puntuación Actividad 1

Autoevaluación de su desempeño para la actividad 1 seleccionada (Navegación por Históricos)

6. Exigencia Mental

Que tan demandante, mentalmente es la tarea

Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

7. Exigencia Física

Que tan demandante, físicamente es la tarea

Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

8. Exigencias Temporales

Que tan fuerte o rápido es el ritmo impuesto para hacer la tarea

Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

9. Rendimiento

Que tan exitoso ha sido para lograr lo que ha requerido

Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

10. Esfuerzo

Que tan duro tiene que trabajar para lograr un adecuado nivel de rendimiento
 Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

11. Nivel de frustración

Que tan inseguro, irritado o estresado y molesto está por la tarea
 Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

Puntuación Actividad 2

Autoevaluación de su desempeño para la actividad 2 seleccionada (Navegación por Complementos)

12. Exigencia Mental

Que tan demandante, mentalmente es la tarea
 Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

13. Exigencia Física

Que tan demandante, físicamente es la tarea
 Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

14. Exigencias Temporales

Que tan fuerte o rápido es el ritmo impuesto para hacer la tarea
 Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

15. Rendimiento

Que tan exitoso ha sido para lograr lo que ha requerido
 Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

16. Esfuerzo

Que tan duro tiene que trabajar para lograr un adecuado nivel de rendimiento
Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

17. Nivel de frustración

Que tan inseguro, irritado o estresado y molesto está por la tarea
Marca solo un óvalo.

	1	2	3	4	5	6	7	8	9	10	
Baja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Alta

Comentarios

Muchas gracias por su valiosa ayuda y tiempo, su opinión es importante para nosotros, ingrésele en la sección de abajo.

18. Ingrese sus comentarios

Con la tecnología de

