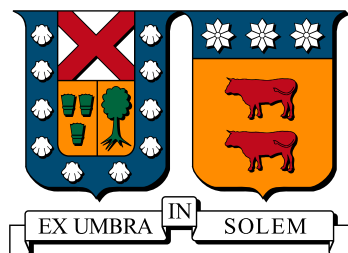


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO – CHILE



**“PROPUESTA DE UN MODELO DE CALIDAD
ENFOCADO EN SOFTWARE INCLUSIVO:
CASO DIALOGAPP”**

ROLANDO EXEQUIEL ROJAS SANFUENTES

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INFORMÁTICO**

PROFESOR GUÍA: PABLO CRUZ

PROFESOR CORREFERENTE: RAÚL MONGE

JULIO 2026

MATERIAL DE REFERENCIA, SU USO NO INVOLUCRA RESPONSABILIDAD DEL AUTOR O DE LA INSTITUCIÓN



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Propuesta de un Modelo de Calidad enfocado en software inclusivo: Caso DialogApp

Nombre del candidato(a): Rolando Exequiel Rojas Sanfuentes

Carrera / Grado: Ingeniería Civil Informática

Campus: Casa Central **Departamento:** Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Pablo Cruz Navea, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

X El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 06/08/2026 **Firma:**

Estudiante o Candidato(a):

Fecha: 06/08/2026 **Firma:**

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Resumen

El desarrollo de aplicaciones inclusivas para la comunidad sorda es un desafío, donde actualmente, las pruebas tradicionales de software no evalúan sus necesidades específicas de comunicación. Para aliviar este problema, el núcleo de esta investigación es proponer un modelo de calidad enfocado en software inclusivo. Este modelo define 8 cualidades (qualities) y para llevar esto a la práctica, se diseñó una Guía de Uso que a través de 24 características observables, busca permitir a los equipos de desarrollo auditar la accesibilidad de sus interfaces de una manera estandarizada.

Para validar la utilidad práctica de la propuesta, se llevó a cabo un caso de estudio en dos fases, solicitando la participación de un panel de evaluadores con experiencia en calidad de software. En primer lugar, se evaluó la aplicación móvil DialogApp usando la guía. Esto permitió diagnosticar el estado del sistema, identificando fortalezas en su navegabilidad e identificando brechas críticas en la naturalidad del avatar 3D. El análisis estadístico de este diagnóstico arrojó un alto nivel de concordancia, mostrando que la pauta logra mitigar en un rango aceptable la subjetividad del evaluador. En segundo lugar, al evaluar el modelo en sí mismo, el análisis cualitativo confirmó que la propuesta es una herramienta clara, completa y actuable para la disciplina. Finalmente, se destacó la necesidad de realizar futuras evaluaciones con equipos interdisciplinarios para cubrir la falta de conocimiento técnico sobre la Lengua de Señas Chilena (LSCh).

Índice

Resumen	II
Índice	III
Índice de cuadros	VI
1. Introducción y definición del problema	1
1.1. Contexto	1
1.2. Planteamiento del problema	2
1.3. Objetivos	3
1.3.1. Objetivo General	3
1.3.2. Objetivos Específicos	4
1.4. Alcance y Limitaciones	4
1.4.1. Alcance	4
1.4.2. Limitaciones	5
1.5. Herramientas de IA y asistencia	5
2. Marco Teórico y Estado del Arte	6
2.1. Marco Conceptual	6
2.1.1. Calidad de Software (ISO/IEC 25010)	6

2.1.2.	Accesibilidad Digital (WCAG)	8
2.1.3.	Discapacidad Auditiva y Lengua de Señas Chilena (LSCh)	10
2.2.	Estado del Arte	11
2.2.1.	Evaluación de Software Inclusivo y de Traducción a Lengua de Señas	11
2.2.2.	Evaluación Heurística en Software Móvil de Salud (mHealth)	12
2.2.3.	Selección de Atributos	14
3.	Propuesta de Solución	16
3.1.	Dimensiones del Modelo	16
3.1.1.	Eje 1: Usabilidad de Interfaz	16
3.1.2.	Eje 2: Calidad del Mensaje	17
3.1.3.	Eje 3: Valor y Aceptación	19
3.2.	Instrumento de Evaluación (Guía de Uso)	19
4.	Validación del Modelo	24
4.1.	Diseño de la Evaluación	24
4.1.1.	Planificación de la Sesión de Evaluación	25
4.1.2.	Análisis de Datos	25
4.2.	Perfil de los Evaluadores	26
4.3.	Presentación del Caso de Estudio: DialogApp	26
5.	Resultados	28
5.1.	Resultados del Instrumento (Guía de Uso)	28
5.1.1.	Hallazgos del Caso de Estudio	28
5.1.2.	Hallazgos estadísticos	31
5.2.	Resultados de Validación	31
5.3.	Análisis Cualitativo y Retroalimentación	33

Conclusiones y Trabajo Futuro	35
Conclusiones Generales	35
Conclusiones de la Validación del Modelo y la Guía de Uso	36
Conclusiones sobre el Caso de Estudio (DialogApp)	36
Trabajo Futuro	37
Bibliografía	39
Anexo 1: Pauta de Evaluación del Software	42
Anexo 2: Pauta de Evaluación del Modelo	47
Anexo 3: Coeficiente de Kendall	50
Coeficiente de Kendall	50
Valor P (p-value)	50
Código	51
Anexo 4: Valores obtenidos en la sesión de Evaluación	52
Evaluación de la Calidad del Software	52
Evaluación del Modelo y Guía de Uso	53

Índice de cuadros

2.1. Tabla de Comparación de Estudios según Atributos y Métodos de Evaluación	13
3.1. Afirmaciones a evaluar del Eje 1: Usabilidad de Interfaz	21
3.2. Afirmaciones a evaluar del Eje 2: Calidad del Mensaje	22
3.3. Afirmaciones a evaluar del Eje 3: Valor y Aceptación	23
5.1. Resultados de Evaluación Promedio por Atributo	29
5.2. Resultados de Validación Promedio por Propiedad	32
A4.1. Valores obtenidos de la Evaluación de la Calidad del Software	52
A4.2. Valores obtenidos de la Evaluación del Modelo y Guía de Uso	54

Capítulo 1

Introducción y definición del problema

1.1. Contexto

En Chile, la comunidad sorda enfrenta barreras de comunicación que limitan su integración social y el ejercicio de sus derechos en igualdad de condiciones. Aunque la normativa nacional reconoce formalmente a la Lengua de Señas Chilena (LSCh) como la lengua oficial de las personas sordas y fomenta su uso [1], existe una gran brecha entre la necesidad real de accesibilidad y los recursos disponibles. Se estima que hay cerca de 180.000 personas con sordera total en el país, mientras que la cantidad de intérpretes certificados ronda apenas los 400 profesionales [2]. Esta escasez crítica dificulta enormemente que las personas sordas puedan interactuar de forma autónoma y oportuna en servicios públicos y privados cotidianos.

Esta situación adquiere un nivel de riesgo mucho mayor en el área de la salud, donde la precisión de la información es vital. La falta de un canal de comunicación directo y fluido entre el paciente sordo y el personal médico obliga, en la gran mayoría de los casos, a depender de familiares o acompañantes para traducir el diálogo. Depender de un tercero, el cual usualmente no está capacitado en terminología médica es un riesgo, además de afectar directamente a la privacidad y la confidencialidad de los datos sensibles del paciente [3]. Además,

este intermediario aumenta la probabilidad de que se omitan detalles importantes sobre los síntomas, dificulta la obtención de un consentimiento informado claro y puede derivar en diagnósticos imprecisos o en el seguimiento incorrecto de los tratamientos, disminuyendo la calidad y seguridad de la atención.

Para ayudar a mitigar esta problemática y muchas otras similares, la tecnología ofrece nuevas alternativas. El avance de la Inteligencia Artificial ha permitido desarrollar sistemas de traducción automática orientados a la accesibilidad. En este escenario surge DialogApp, un proyecto que sale de Taller de Desarrollo de Proyectos de Informática, una asignatura de quinto año de la carrera de Ingeniería Civil Informática en la UTFSM donde se busca obtener una solución informática para un problema real mediante la aplicación de los conocimientos adquiridos durante la carrera. La asignatura culmina con la presentación de los resultados en la Feria de Software, una instancia que es abierta a toda la comunidad. DialogApp consiste en una aplicación móvil diseñada para funcionar como un traductor bidireccional en tiempo real entre el español y la LSCh. Siguiendo con el ejemplo del ámbito médico, por un lado, el sistema utiliza la cámara del dispositivo para capturar e interpretar las señas del paciente y traducirlas al español; por otro, se apoya en un avatar 3D para emitir las respuestas del médico hacia la lengua de señas. Su objetivo principal es ofrecer una herramienta accesible desde un teléfono convencional, facilitando una comunicación independiente en la consulta médica. Sin embargo, la creación de este tipo de soluciones introduce una interrogante crítica: asegurar que tanto la captura de las señas como su representación a través de un avatar sean precisas, comprensibles y útiles para el paciente requiere de criterios de evaluación específicos, un aspecto que actualmente representa un desafío metodológico para la Ingeniería de Software.

1.2. Planteamiento del problema

El desarrollo de tecnologías de accesibilidad, como los sistemas de comunicación mediante avatares de lengua de señas, avanza rápidamente. Sin embargo, lograr que la comunidad sorda adopte estas herramientas sigue siendo un desafío [4]. Un software de traducción no

es útil en la práctica si su diseño no considera los modelos mentales y los requerimientos de interacción de los usuarios finales. Las evaluaciones indican que la precisión técnica de la traducción es insuficiente si el avatar no tiene fluidez, expresiones faciales realistas y movimientos naturales [5, 6].

Actualmente, la Ingeniería de Software utiliza estándares como la norma ISO/IEC 25010 para evaluar productos. Sin embargo, estos modelos de calidad generales son insuficientes para medir aplicaciones orientadas a personas sordas. Las métricas automáticas tradicionales, como BLEU o el índice de error de señas, no capturan la complejidad de la lengua de señas, ya que omiten factores como los marcadores gramaticales no manuales y el uso del espacio tridimensional [7, 8]. Por esto, la literatura indica que la validación de estas tecnologías debe centrarse en el usuario [8]. Variables como la edad de adquisición de la lengua y el nivel de fluidez afectan directamente cómo se perciben y aceptan los avatares generados por computadora [6].

La falta de un marco de evaluación adaptado representa un riesgo en el ciclo de vida del software. La ausencia de heurísticas de usabilidad para personas sordas [9] y la falta de herramientas estandarizadas para evaluar aplicaciones móviles de inclusión [10] pueden resultar en la liberación de productos poco usables. Esto genera frustración, rechazo y desconfianza sobre la privacidad de los datos [3]. Por lo tanto, es necesario proponer un modelo de calidad y una guía de uso específica para software inclusivo. Este modelo debe definir cualidades (*qualities*) como la precisión gramatical de la seña, la legibilidad, la adaptabilidad cultural y la usabilidad [5], para poder proponer una guía de uso en base a características observables que permitan validar de forma estandarizada aplicaciones como DialogApp.

1.3. Objetivos

1.3.1. Objetivo General

- Proponer un modelo que permita la evaluación de la calidad del software enfocado en inclusión y accesibilidad para la población en situación de discapacidad auditiva.

1.3.2. Objetivos Específicos

- Analizar y seleccionar las necesidades cualificables de la población en situación de discapacidad auditiva en materia de software.
- Parametrizar los criterios de Calidad obtenidos de la población sorda en un modelo para software inclusivo enfocado en personas en situación de discapacidad.
- Validar y analizar la aplicación DialogApp con el modelo propuesto.

1.4. Alcance y Limitaciones

1.4.1. Alcance

El alcance de este trabajo se centra exclusivamente en el diseño y validación teórica-práctica de una propuesta integral conformada por un modelo de calidad de software junto a un instrumento de evaluación orientado a aplicaciones móviles inclusivas para la comunidad sorda.

El trabajo abarca:

- La identificación y selección de atributos de calidad pertinentes derivados de estándares internacionales (tales como la norma ISO/IEC 25010) [11] y directrices de accesibilidad (como WCAG) [12], adaptados a las necesidades de interacción de usuarios con discapacidad auditiva.
- La estructuración de un modelo de calidad que proponga dichas necesidades en cualidades, junto a un instrumento de evaluación (Guía de Uso) que permita estandarizar la aplicación del modelo a un software inclusivo mediante métricas y características observables.
- La validación de la propuesta mediante un caso de estudio aplicado a una aplicación (DialogApp), utilizando un enfoque de auditoría con la Guía de Uso, para posteriormente realizar una validación directa del modelo mediante una pauta de evaluación.

1.4.2. Limitaciones

Para asegurar que el proyecto sea viable y se cumpla dentro de los plazos establecidos, se han definido ciertas limitaciones. En primer lugar, el modelo de calidad y su respectiva Guía de Uso se enfocan exclusivamente en las necesidades de accesibilidad de las personas con discapacidad auditiva y en el uso de la Lengua de Señas Chilena (LSCh), por lo que no abordará cualidades ni características para otras discapacidades, como visuales, motoras o cognitivas. Además, dado el estado actual de prototipo de la aplicación DialogApp y el tiempo disponible, no se realizarán pruebas masivas ni focus groups con usuarios finales. En su lugar, la validación dependerá de una auditoría realizada por evaluadores, enfocada en comprobar la objetividad y la evidencia de que la Guía de Uso es actuable. Finalmente, al tratarse de un trabajo centrado en el aseguramiento de la calidad y la evaluación, el proyecto no contempla programar, modificar el código fuente ni agregar nuevas funcionalidades al software estudiado.

1.5. Herramientas de IA y asistencia

Durante la redacción y desarrollo de esta memoria se utilizó el modelo de lenguaje grande (LLM) Gemini como herramienta de asistencia de redacción, ortografía y código. Su uso se destinó a mejoras en la redacción de los capítulos, corrección ortográfica, y apoyo en la generación de código (script de análisis estadístico).

Capítulo 2

Marco Teórico y Estado del Arte

2.1. Marco Conceptual

2.1.1. Calidad de Software (ISO/IEC 25010)

En el ámbito de la Ingeniería de Software, la calidad del producto se define formalmente como el grado en que un sistema informático cumple con los requisitos especificados y satisface las expectativas de sus usuarios finales [13]. Para estandarizar su medición, la Organización Internacional de Normalización y la Comisión Electrotécnica Internacional establecen el modelo de referencia ISO/IEC 25010. En su actualización del año 2023, este estándar proporciona un marco estructurado para especificar, medir y evaluar las cualidades del producto a lo largo de su ciclo de vida, agrupando la calidad en nueve cualidades principales [11]:

- **Adecuación Funcional** (Functional suitability): Capacidad del producto para proporcionar funciones que satisfacen las necesidades declaradas e implícitas bajo condiciones de uso específicas.
- **Eficiencia de Desempeño** (Performance efficiency): Capacidad para realizar funciones de manera óptima respecto a los tiempos de respuesta y el uso eficiente de recursos.

- **Compatibilidad** (Compatibility): Capacidad del sistema para intercambiar información y operar en un entorno compartido con otros productos sin causar impactos perjudiciales.
- **Capacidad de Interacción** (Interaction capability): (antes llamada Usabilidad). Grado en que los usuarios pueden interactuar con el sistema y completar tareas a través de la interfaz de manera efectiva.
- **Fiabilidad** (Reliability): Capacidad de operar sin interrupciones ni fallos durante un período de tiempo determinado.
- **Seguridad** (Security): Capacidad para proteger la información y los datos, garantizando niveles de acceso autorizados y defensa contra ataques.
- **Mantenibilidad** (Maintainability): Capacidad del sistema para ser modificado, corregido, mejorado o adaptado de forma eficaz y eficiente.
- **Flexibilidad** (Flexibility): (antes denominada Portabilidad). Capacidad del producto para adaptarse a cambios en sus requisitos, contextos de uso o en el entorno del sistema.
- **Seguridad Física u Operacional** (Safety): Cualidad introducida en 2023 que define la capacidad para evitar estados que pongan en peligro la vida humana, la salud, la propiedad o el medio ambiente.

Si bien este modelo proporciona una base general, en el contexto del desarrollo de aplicaciones móviles interactivas orientadas a la salud (mHealth), la literatura destaca tres dimensiones específicas. La Adecuación Funcional es necesaria desde la fase de requisitos para asegurar que la aplicación entregue resultados precisos y alineados con su propósito clínico [13]. Por su parte, la Capacidad de Interacción exige considerar las dimensiones limitadas de las pantallas para minimizar errores de manipulación [13]. Finalmente, la Compatibilidad es necesaria para garantizar una operabilidad cruzada fluida entre múltiples plataformas y sistemas operativos móviles [13].

A pesar de la solidez general de la norma ISO/IEC 25010, la revisión de la literatura advierte que los modelos de calidad puramente genéricos presentan limitaciones al momento de evaluar software enfocado en la accesibilidad [14, 15]. En primer lugar, estas pautas suelen basarse en principios abstractos que no siempre cubren las necesidades cognitivas y culturales únicas de poblaciones diversas, lo que dificulta la evaluación de sistemas con alto contenido multimedia [14].

Específicamente en el desarrollo de avatares traductores a lengua de señas, la norma no puede aplicarse de manera estricta o automatizada. La complejidad lingüística y la diversidad de estilos de la lengua impiden el uso de métricas estandarizadas de software, donde atributos críticos como la naturalidad del movimiento corporal, la legibilidad y la plausibilidad de las expresiones faciales necesitan de una evaluación humana directa por parte de usuarios nativos [5]. Esta dificultad de uso de las cualidades de la ISO 25010 frente a barreras de interacción reales hace necesario ir más allá de los modelos generales, haciendo necesaria la formulación de pautas de evaluación heurística adaptadas específicamente a la población objetivo [15, 9].

2.1.2. Accesibilidad Digital (WCAG)

La accesibilidad digital se define como el grado en que un producto tecnológico puede ser utilizado por usuarios con discapacidades específicas para alcanzar sus objetivos con efectividad, eficiencia y satisfacción [15]. En el contexto del desarrollo de software contemporáneo, este concepto ha evolucionado hacia el modelo de Diseño Inclusivo Responsivo (RiD, por sus siglas en inglés). Este enfoque sostiene que la tecnología debe adaptarse de manera integral a la independencia funcional e intelectual del usuario durante todo el ciclo de vida del producto, pasando de la “adaptación al dispositivo” para centrarse en la “adaptación al usuario” [16]. Para la población con discapacidad auditiva, la unión entre las barreras de accesibilidad y los problemas de usabilidad general se conceptualizan bajo el término DWUX (Deaf Web User Experience), el cual evalúa tanto la utilidad práctica como la satisfacción del usuario al interactuar con interfaces digitales [15].

A nivel global, el principal instrumento normativo para la construcción de interfaces accesibles son las Pautas de Accesibilidad para el Contenido Web (WCAG) desarrolladas por W3C. Sin embargo, en la práctica, la aplicación de este estándar se centra principalmente en proveer alternativas en forma de texto como subtítulos o transcripciones para el contenido multimedia, las cuales representan el cumplimiento base (Nivel A) [17, 18]. Por otro lado, entregar interfaces de interacción directa en lengua de señas es categorizada por las WCAG como un requisito de Nivel AAA [17], por lo que en la práctica, esto se traduce en que la lengua de señas es considerada una característica opcional y poco frecuente como un estándar de cumplimiento exigible [18].

La literatura critica bastante esta postergación normativa, señalando que la suposición de que el texto escrito equivale a accesibilidad es un error fundamental de diseño. La comunidad sorda, en especial las personas prelinguales (personas que pierden la capacidad auditiva previo a desarrollar el habla y el lenguaje), experimenta severas barreras de alfabetización en el idioma oral-escrito dominante [15]. Se estima que un alto porcentaje de esta población enfrenta dificultades críticas para comprender información escrita compleja, operando frecuentemente con un nivel de comprensión lectora equivalente a la educación primaria [4]. Esto ocurre porque las lenguas de señas poseen una estructura morfológica y una sintaxis espacial propia, convirtiendo al texto escrito, esencialmente, en una segunda lengua [17, 18]. Esto significa que depender de normativas genéricas basadas exclusivamente en texto induce a la exclusión digital y a una mala experiencia en servicios críticos como la atención clínica [4, 18].

Para disminuir esta brecha y garantizar una autonomía real, es necesario proveer servicios digitales directamente en la lengua materna del usuario [8]. Ante la escasez de intérpretes humanos, sus altos costos y la vulneración de la privacidad que suponen en citas médicas, la integración de avatares de lengua de señas (virtual signers) se posiciona como una alternativa viable para ampliar el acceso a la información [6, 4]. Sin embargo, estos sistemas pueden ser rechazados si su calidad no se evalúa de forma rigurosa. La generación de avatares suele carecer de fluidez natural y omitir marcadores no manuales esenciales (expresiones faciales y

movimientos corporales), provocando que la interfaz caiga en el “valle inquietante” (uncanny valley) y resulte incomprensible [5, 19]. Por lo tanto, validar estas herramientas requiere definir características medibles para la exactitud gramatical y la adaptabilidad cultural con un enfoque de diseño participativo que involucre directamente a la propia comunidad [8, 18].

2.1.3. Discapacidad Auditiva y Lengua de Señas Chilena (LSCh)

El diseño de software para la comunidad sorda requiere comprender que su modelo cognitivo es distinto al procesamiento auditivo de las personas oyentes. Los usuarios sordos estructuran la información de manera visual y espacial, procesando la interacción en un formato de cuatro dimensiones que abarca el uso del espacio 3D y el tiempo [17, 15]. En este marco, lenguas como la Lengua de Señas Chilena (LSCh) transmiten información coordinando al mismo tiempo las manos, la postura y la mirada. Un factor crítico en esta estructura espacial es la dependencia de marcadores gramaticales no manuales, tales como micro-expresiones faciales y movimientos corporales. Estos elementos no actúan solo como meros complementos, sino como elementos sintácticos indispensables para estructurar el mensaje y evitar ambigüedades del lenguaje [5, 19].

Esta naturaleza propia de las lenguas de señas trae desafíos técnicos para la Ingeniería de Software orientada a la generación de avatares y sistemas de visión computacional. Tecnológicamente, la dificultad de replicar la “gramática facial” simultánea y transiciones fluidas produce representaciones robóticas que caen en el “valle inquietante” (uncanny valley), generando rechazo e incomprensión en el usuario final [6, 8]. Adicionalmente, la captura y procesamiento de la lengua espacial enfrenta barreras físicas como la oclusión entre extremidades, la degradación de imagen por velocidad (motion blur) y la pérdida de contexto dinámico al intentar forzar información tridimensional en modelos de datos de texto lineal [7, 5]. Por lo que, la guía de uso de un modelo de evaluación de calidad para este tipo de software debe tener en cuenta características medibles que prioricen la fidelidad de esta representación tridimensional y simultánea por sobre la simple precisión gramatical en texto.

2.2. Estado del Arte

2.2.1. Evaluación de Software Inclusivo y de Traducción a Lengua de Señas

La evaluación de aplicaciones móviles orientadas a la comunidad sorda requiere un enfoque dual: la validación de la interfaz gráfica general y la medición específica del motor de traducción tridimensional. A nivel de interfaz, la literatura reciente adapta las cualidades clásicas de calidad para mitigar las barreras cognitivas y sensoriales de los usuarios prelinguales. En términos de Capacidad de Interacción y Usabilidad, autores como Yeratziotis y Zaphiris [15] y Xiong [9] priorizan métricas de carga cognitiva visual reducida, alertas hápticas (vibración) y alternativas de lectura fluida frente a la baja alfabetización textual. Para validar estas cualidades, proponen inspecciones mediante pautas heurísticas adaptadas, tales como HE4DWUX (Heuristic Evaluation for Deaf Web User Experience) y HHITA (Heuristics for the Hearing Impaired Training APP). Además, para asegurar la Adecuación Funcional y Capacidad de Aprendizaje, se evalúa la presencia de retroalimentación visual inmediata y el uso de iconografía de alta facilidad de reconocer, metodologías que son testeadas mediante protocolos adaptados como el “Think by Signs” y checklists derivados de la norma ISO/IEC 25010 [17, 13].

Mientras tanto, la evaluación a nivel de traducción (avatares 3D e IA) difiere del software tradicional. Existe un consenso en la literatura en que las métricas automatizadas de traducción unidimensional (como BLEU, WER o NIST) tienen limitaciones al evaluar lengua de señas [7, 19]. Esto ocurre porque estas fórmulas penalizan alteraciones gramaticales visuales correctas y no procesan de manera adecuada las dimensiones del espacio y tiempo del idioma, provocando una pérdida de información semántica [5].

Ante las limitaciones de estas métricas, la literatura define la calidad de los avatares mediante un panel de evaluación estructurado en tres características principales. Primero, la Exactitud Gramatical, medida a través de la posición de la palma junto a los dedos y la aplicación de

predicados clasificadores espaciales las cuales son revisadas por los evaluadores mediante cuestionarios objetivos de coincidencia visual [7, 5]. Segundo, la Naturalidad y Experiencia de Usuario, que cuantifica la sincronización de expresiones faciales y de boca (marcadores no manuales) y la fluidez para reducir el rechazo por el efecto de valle inquietante (uncanny valley). Estas métricas se validan utilizando el índice NASA-TLX para medir la carga mental y cuestionarios UEQ-S para la calidad pragmática [8, 19]. Finalmente, la Adaptabilidad Cultural, que evalúa la confianza y el contexto en el que está el avatar empleando marcos teóricos como el modelo TAMSA (Technology Acceptance Model of Signing Avatars) [18]. La Tabla 2.1 es un resumen de las cualidades, métricas y métodos de evaluación observados del estado del arte, los cuales sirven como base teórica para la formulación del modelo de calidad.

2.2.2. Evaluación Heurística en Software Móvil de Salud (mHealth)

En el ámbito de la salud móvil (mHealth), medir la calidad de una aplicación basándose únicamente en las estrellas o reseñas de las tiendas virtuales resulta insuficiente y subjetivo [20]. Para asegurar que las herramientas médicas digitales sean confiables y efectivas, la comunidad científica utiliza instrumentos estandarizados. Uno bien reconocido es la Escala de Calificación de Aplicaciones Móviles (MARS, por sus siglas en inglés), la cual clasifica el software mediante cuatro dimensiones objetivas: el nivel de participación (*engagement*), la funcionalidad, la estética y la calidad de la información [10].

Un requisito de instrumentos como MARS es que no están diseñados para el usuario común, sino que exigen ser aplicados por un equipo multidisciplinario, abarcando profesionales de la salud, diseñadores e informáticos [10]. La literatura respalda este enfoque, conocido como evaluación heurística. Se considera un método ágil y rentable que permite detectar problemas de usabilidad desde las primeras fases del desarrollo, ahorrando recursos valiosos antes de requerir pruebas masivas con pacientes [15].

A pesar de contar con estos estándares de evaluación médica, el mercado actual de aplicaciones para personas sordas presenta altas tasas de abandono debido a diseños deficientes

Cuadro 2.1: Tabla de Comparación de Estudios según Atributos y Métodos de Evaluación

Autor(es) y Año	Dominio	Cualidad Principal	Métricas Específicas	Método de Evaluación
Yeratziotis y Zaphiris, 2018	Interfaz	Usabilidad e Interacción	Alternativas visuales al audio, lectura fluida, navegación intuitiva	Evaluación Heurística (HE4DWUX) y encuestas de usuarios
Xiong et al., 2020	Interfaz	Usabilidad y Prevención de Errores	Alertas de vibración, diseño minimalista, retroalimentación visual inmediata	Testeo del Modelo Heurístico (HHITA)
Bianchini et al., 2017	Interfaz	Capacidad de Aprendizaje	Minimización de información textual, diseño basado en alta iconicidad	Protocolo “Think by Signs” y cuestionario de satisfacción QUIS
Moumane et al., 2024	Interfaz	Adecuación Funcional	Interoperabilidad, control de datos del usuario, acciones operativas claras	Checklists basados en ISO/IEC 25010 e ISO/IEC 25023
Zhao et al., 2024	Avatar	Exactitud Gramatical	Posición de dedos, orientación de palma, marcadores sintácticos de seña	Evaluación humana objetiva (cuestionarios de opción múltiple y Likert)
Huenerfauth et al., 2008	Avatar	Comprensibilidad	Reconocimiento de clasificadores y parámetros espaciales 3D	Tareas de comprensión de coincidencia visual-espacial humana
Wasserroth et al., 2025	Avatar	Naturalidad y UX	Marcadores no manuales (mouthings), efecto valle inquietante, fluidez	Cuestionarios NASA-TLX (estrés/esfuerzo) y UEQ-S (experiencia de usuario)
Othman et al., 2024	Avatar	Aceptación y Adaptabilidad Cultural	Confianza, utilidad percibida, indumentaria/cultura local del avatar	Modelo TAMSA con entrevistas semiestructuradas
Bennbaia, 2022	Avatar	Aceptabilidad	Movimiento corporal natural, expresiones emocionales coherentes	Focus Groups con la comunidad sorda y métrica SUS

[20]. Como se discutió en secciones anteriores, depender de pautas genéricas orientadas a la transcripción de texto resulta excluyente para usuarios prelocutivos [18]. Para resolver esta brecha, los investigadores subrayan la necesidad estricta de tomar escalas como MARS y modificarlas, añadiendo criterios específicos para la sordera. Esto significa evaluar variables que los modelos tradicionales ignoran, tales como la correcta delimitación visual del “espacio de señas”, la eliminación de distracciones visuales en la interfaz y la pertinencia de las funciones de asistencia [20]. Esta necesidad de adaptar los instrumentos de evaluación a las características del usuario no solo aplica a la discapacidad auditiva, sino que incluso es un desafío en el desarrollo de software para la comunidad neurodivergente [14].

2.2.3. Selección de Atributos

La revisión de la literatura evidencia que los estándares de calidad generales no son suficientes para evaluar software dirigido a la comunidad sorda. La norma ISO/IEC 25010 resulta general para este contexto, las directrices de accesibilidad (WCAG) se enfocan principalmente en el texto (excluyendo a quienes tienen la lengua de señas como primer idioma) y las escalas médicas como MARS requieren adaptaciones para evaluar correctamente la interacción visual. Para construir el modelo de calidad, se seleccionaron las cualidades relevantes que identificaron de estos estudios. Estas cualidades se agrupan en tres ejes que adaptan y especializan la norma ISO/IEC 25010:

- **Eje 1: Usabilidad de Interfaz** (Basado en Capacidad de Interacción): Va más allá de ofrecer subtítulos. Se enfoca en que la interfaz sea usable para una persona sorda mediante una navegación fluida, retroalimentación inmediata (alertas visuales o por vibración), baja sobrecarga visual y el uso de íconos claros que eviten depender de la lectura de textos largos.
- **Eje 2: Calidad del Mensaje** (Basado en Adecuación Funcional): Descarta las métricas automáticas (como BLEU) que tienen limitaciones con la lengua de señas y prioriza la evaluación humana. Considera si el avatar es gramaticalmente correcto (uso del espacio y las manos) y si sus movimientos son naturales. Esto incluye evaluar sus expresiones faciales y corporales para que el diseño sea más natural.

- **Eje 3: Valor y Aceptación** (Basado en estudios mHealth): Incorpora la evaluación mediante evaluadores en accesibilidad. Considera si el usuario confía en el sistema, si el avatar se adapta a su cultura local, y si la información médica que entrega la aplicación es segura y útil en un entorno clínico real.

A partir de cada uno de estos tres ejes se definirán las cualidades del modelo de calidad. En el siguiente capítulo, se describirán las cualidades de cada eje, donde a posterior estas cualidades se estandarizarán en características medibles en una guía de uso, diseñada para ser aplicada por un panel de evaluadores.

Capítulo 3

Propuesta de Solución

3.1. Dimensiones del Modelo

Como se vió en la revisión del estado del arte, evaluar un software diseñado para la comunidad sorda requiere tomar en cuenta factores que los estándares tradicionales no suelen contemplar. Para estructurar la información extraída de la literatura, esta se organizó en tres ejes de evaluación.

El propósito de esta división es separar las distintas partes de la aplicación: el funcionamiento de la interfaz no debe confundirse con la calidad de la traducción del avatar, ni con el valor que la herramienta entrega en su contexto de uso. De cada uno de estos tres ejes generales se desglosan subcategorías más específicas, denominadas cualidades (qualities). Estas cualidades son los conceptos centrales del modelo, donde posteriormente se llevarán estos conceptos a características medibles en una guía de uso.

3.1.1. Eje 1: Usabilidad de Interfaz

Este eje aborda la interfaz de la aplicación. Su propósito es analizar la capa de presentación y navegación, asegurando que el software se adapte a los usuarios con discapacidad auditiva y reduzca la barrera que representa la lectura.

Claridad Visual

Se enfoca en la interfaz gráfica en función de su carga cognitiva. Dado que los usuarios prelocutivos procesan la información de forma visual, la saturación de elementos en pantalla dificulta el uso. Las características a observar incluyen: el uso de espacios en blanco, la aplicación de alto contraste para destacar elementos interactivos y la ausencia de bloques extensos de texto. Se evita la dependencia de instrucciones escritas, prefiriendo un diseño minimalista.

Navegabilidad Intuitiva

Representa la facilidad con la que el usuario utiliza la aplicación sin aprendizaje previo. Se deben considerar la longitud de los flujos de tareas (minimizando la cantidad de interacciones para alcanzar un objetivo) y el uso de iconos claros. Estos iconos deben representar acciones de manera universal, logrando que la navegación se base en el reconocimiento visual y no en la lectura de menús.

Tolerancia a Errores

Representa la capacidad del sistema para prevenir fallos y guiar al usuario en su recuperación. En este contexto, los mensajes de error en texto son ineficaces. Se debe considerar la presencia de retroalimentación sensorial inmediata, como la integración de alertas hápticas (vibración) y cambios visuales evidentes (colores o iconos de detención) ante acciones incorrectas o no deseadas.

3.1.2. Eje 2: Calidad del Mensaje

Este eje aborda el contenido dinámico del software que consiste en la sección de traducción, donde se debe enfocar en la fidelidad de los conceptos y biomecánica de la lengua de señas generada, como alternativa a las limitaciones de las evaluaciones de traducción basadas

únicamente en texto.

Precisión Espacial

Representa la exactitud esperada de la representación tridimensional. Dado que la lengua de señas utiliza el espacio físico con fines gramaticales, las características relevantes son la configuración de las manos (posición de manos y dedos), la orientación de las palmas respecto al cuerpo, y el uso del “espacio de señas”.

Naturalidad

Considera el realismo biomecánico del modelo 3D para mitigar el efecto del valle inquietante (uncanny valley), ya que un avatar con movimientos precisos pero rígidos no logra una comunicación efectiva. Se enfoca en la fluidez durante la transición entre señas junto con la sincronización de expresiones faciales y de boca (marcadores no manuales). Esto implica que el avatar utilice microexpresiones faciales, patrones labiales (mouthings) y movimientos de cabeza que tengan sentido con el significado de la seña.

Comprensibilidad

Corresponde a la claridad del mensaje emitido y la adaptación del sistema al usuario. Se debe considerar la capacidad del avatar para diferenciar señas que comparten configuraciones manuales similares, pero con significados distintos (un ejemplo es la letra Ñ que es tiene la misma configuración de la N pero con un movimiento horizontal). Además, se consideran controles de interacción sobre la animación, como la opción de repetir, pausar o ajustar la velocidad del gesto sin perder la fluidez.

3.1.3. Eje 3: Valor y Aceptación

Este eje aborda el “propósito” de la aplicación, el de conectar el software con el entorno y propósito específico, para definir si el software cumple con los estándares éticos, de seguridad y de pertenencia cultural necesarios para su adopción por parte de la comunidad.

Fiabilidad (de la Información)

Representa la exactitud de la herramienta dentro de su contexto de operación, considerando que una traducción ambigua en aplicaciones críticas puede generar un impacto en el usuario. Se debe considerar la precisión del avatar al traducir terminología técnica o especializada (como conceptos médicos, legales o académicos), y evitar las traducciones literales que alteran el sentido original del mensaje. Además, se analiza la seguridad de la interfaz para comprobar la privacidad en el uso de datos sensibles.

Aceptación Cultural

Incluye la confianza, el sentido de pertenencia y la empatía que la herramienta genera en la comunidad. Se enfoca en verificar el uso de señas, modismos y reglas de la lengua local (para este caso de estudio, la Lengua de Señas Chilena - LSCh), restringiendo la mezcla con lenguas extranjeras o de otros dialectos. Además, el diseño estético y la apariencia del avatar deben proyectar un nivel de profesionalismo acorde al propósito de la aplicación, para aumentar la confianza del usuario y reducir las tasas de abandono.

3.2. Instrumento de Evaluación (Guía de Uso)

Como ya se tienen definidas las cualidades del modelo, el siguiente paso es diseñar una guía de uso (o método de evaluación), y basándonos en lo revisado en el Estado del Arte se optó por el diseño de un instrumento de evaluación basado en una pauta heurística. La idea de esta guía de uso es transformar la teoría en una herramienta estandarizada que permita a los

evaluadores analizar el software de manera objetiva.

Para su construcción, se optó por un formato de afirmaciones evaluables mediante una escala Likert que va desde 1 a 5 puntos, donde 1 significa “Totalmente en desacuerdo” y 5 representa “Totalmente de acuerdo”.

Se estructuró la guía de uso asignando tres afirmaciones por cada cualidad del modelo, y dado que el modelo cuenta con 8 cualidades, la guía se compone de 24 puntos de evaluación, cantidad que busca un equilibrio entre la profundidad del análisis y la carga cognitiva del evaluador. La idea es que con 24 la pauta sea lo suficientemente exhaustiva para detectar defectos, pero a la vez ágil para mantener la concentración del evaluador.

En los cuadros 3.1, 3.2 y 3.3 se ven las afirmaciones a evaluar en la guía de uso propuesta, están separadas según la cualidad y cada cuadro representa cada uno de los tres ejes propuestos.

Cuadro 3.1: Afirmaciones a evaluar del Eje 1: Usabilidad de Interfaz

Cualidad	ID	Afirmación a evaluar
Claridad Visual	1.1a	La interfaz usa un mínimo de textos largos y utiliza los espacios en blanco adecuadamente para no saturar al usuario.
	1.1b	Existe un contraste adecuado entre los elementos interactivos y el fondo que facilita su visualización y lectura rápida.
	1.1c	La pantalla está libre de elementos decorativos o animaciones que puedan generar distracción visual o sobrecarga cognitiva.
Navegabilidad Intuitiva	1.2a	La aplicación utiliza iconografía estandarizada y clara que permite navegar sin depender de menús textuales complejos.
	1.2b	Los flujos para completar tareas son directos y requieren un número mínimo de interacciones (clics o toques de pantalla).
	1.2c	La ubicación de los botones y controles principales se mantiene consistente en todas las pantallas, reduciendo el esfuerzo de aprendizaje.
Tolerancia a Errores	1.3a	El sistema emite respuestas visuales o de vibración (hápticas) claras e inmediatas cuando el usuario realiza una acción.
	1.3b	Ante un error, el sistema guía al usuario de manera visual, evitando entregar únicamente mensajes de fallo en formato de texto.
	1.3c	La interfaz ofrece la opción de cancelar o deshacer fácilmente una acción crítica antes de que se ejecute de manera definitiva.

Cuadro 3.2: Afirmaciones a evaluar del Eje 2: Calidad del Mensaje

Cualidad	ID	Afirmación a evaluar
Precisión Espacial	2.1a	La configuración de las manos y la orientación de las palmas son gramaticalmente correctas respecto a la seña emitida.
	2.1b	El emisor del mensaje utiliza correctamente el espacio tridimensional para situar sujetos y acciones dentro de la oración.
	2.1c	Las trayectorias de movimiento entre un punto espacial y otro son exactas y no generan deformaciones anatómicas en la animación.
Naturalidad	2.2a	Los movimientos corporales y la transición entre las distintas señas se perciben fluidos y evitan la rigidez robótica.
	2.2b	Se sincronizan adecuadamente las expresiones faciales y corporales con el significado de la seña.
	2.2c	El ritmo de la traducción y las pausas entre oraciones imitan la cadencia y respiración de una conversación humana real.
Comprensibilidad	2.3a	El sistema permite diferenciar con claridad visual aquellas señas que tienen movimientos o configuraciones muy similares.
	2.3b	La interfaz ofrece herramientas para repetir el mensaje o ajustar su velocidad sin que se pierda la legibilidad de la seña.
	2.3c	El mensaje conserva su intención original sin perder significado durante la traducción.

Cuadro 3.3: Afirmaciones a evaluar del Eje 3: Valor y Aceptación

Cualidad	ID	Afirmación a evaluar
Fiabilidad de la Información	3.1a	El sistema traduce con precisión y sin ambigüedades la terminología técnica o crítica propia del contexto de la aplicación.
	3.1b	En los flujos que gestionan datos sensibles, se transmite al usuario seguridad y claridad respecto a la protección de su privacidad.
	3.1c	La herramienta es consistente y no entrega traducciones contradictorias al interpretar un mismo concepto en distintas pantallas.
Aceptación Cultural	3.2a	La traducción respeta fielmente la gramática y el vocabulario local, sin mezclar modismos extranjeros.
	3.2b	El diseño estético general transmite el profesionalismo adecuado para el dominio de la aplicación, generando confianza en el usuario.
	3.2c	La apariencia del avatar resulta empática y evita el uso de estereotipos o elementos visuales que puedan resultar extraños para la comunidad local.

Capítulo 4

Validación del Modelo

4.1. Diseño de la Evaluación

El objetivo principal de esta sección es establecer la metodología para validar la propuesta de modelo de calidad junto con la guía de uso, donde el foco irá en 2 puntos: el primer foco es evaluar el producto (DialogApp) mediante la guía propuesta para buscar evidencia tanto cuantitativa (buscando el nivel de concordancia entre evaluadores) como cualitativa de que la guía de uso es actuable y apropiada para evaluar. En segundo lugar, se buscará evaluar el modelo junto a la guía de uso en sí mismos, para recolectar evidencia que permita validar la utilidad y correctitud de la propuesta. Para abarcar estos dos focos, se usarán dos instrumentos de forma secuencial:

- **Instrumento 1** (Evaluación de la Calidad del Software): Corresponde a la aplicación práctica del modelo a través de la Guía de Uso de 24 puntos diseñada en el Capítulo 3. Los evaluadores utilizarán esta herramienta para evaluar a DialogApp, calificando cada afirmación con una nota del 1 al 5. Ver en **Anexo 1**.
- **Instrumento 2** (Evaluación del Modelo y Guía de Uso): Se entrega inmediatamente después de usar el primer instrumento. Su objetivo es evaluar el modelo y la guía de uso como tales. Mide qué tan fáciles fueron de aplicar, si las instrucciones fueron

claras y si abarcan todos los aspectos necesarios para auditar software inclusivo. Ver en **Anexo 2**.

4.1.1. Planificación de la Sesión de Evaluación

Para que los resultados sean comparables, las evaluaciones se realizarán de manera individual y cada sesión dura hasta 45 minutos, dividiéndose en tres pasos:

1. **Contextualización** (5 min): Se le entrega al evaluador el acceso a DialogApp y el Instrumento 1. Se le explica su rol y un pequeño contexto sobre la aplicación y el modelo propuesto de calidad.
2. **Evaluación del Software** (hasta 30 min): El evaluador revisa y evalúa el software, donde se le sugieren 3 tareas que cubren los flujos principales de la aplicación. Mientras interactúa con el sistema, el evaluador califica las afirmaciones del Instrumento 1 sin ayuda del investigador.
3. **Evaluación del Modelo y Guía de Uso** (hasta 10 min): Al terminar la revisión de la aplicación, se retira el Instrumento 1 y se le entrega el Instrumento 2. En este paso, el evaluador opina sobre su experiencia utilizando el modelo y guía de uso propuestos.

4.1.2. Análisis de Datos

El análisis de los resultados es el núcleo de la validación del modelo y se divide en dos enfoques:

- **Análisis del Instrumento 1** (Coeficiente de Kendall): Los puntajes del 1 al 5 obtenidos en el Instrumento 1 se analizarán usando el Coeficiente de Concordancia de Kendall (W). Este coeficiente mide el nivel de acuerdo entre los distintos evaluadores. Si el cálculo muestra un acuerdo fuerte (W cercano a 1), se puede argumentar que el modelo es adecuado. Si hay desacuerdo, el resultado puede ser útil para identificar posibles problemas a mejorar en el futuro.

- **Análisis del Instrumento 2:** Las respuestas en la escala de 1 a 5 del segundo instrumento se tabularán para confirmar si los evaluadores consideran que el modelo es una herramienta útil para su trabajo profesional. Además, las preguntas abiertas entregarán información clave para proponer futuras versiones del modelo.

4.2. Perfil de los Evaluadores

Para que los resultados de la validación sean rigurosos y que el análisis tenga validez, es fundamental definir quiénes realizarán la evaluación. En esta ocasión, se determinó conformar un panel de 4 evaluadores los cuales cumplen con el siguiente perfil:

- **Formación Técnica:** Profesionales de Ingeniería Civil Informática, Ingeniería en Informática o carreras afines, puesto que deben tener la capacidad de analizar el software desde su construcción.
- **Conocimientos:** Experiencia o formación previa en Aseguramiento de la Calidad (QA), diseño de experiencia de usuario (UX/UI) o Interacción Humano Computador.
- **Conocimiento de Inclusión:** No es requisito que los evaluadores dominen la Lengua de Señas Chilena (LSCh). Dado que hay una sección de contextualización, se les entregará lo necesario para poder interactuar con la aplicación, además, el propio modelo y guía deben ser lo suficientemente claros como para guiarlos a identificar si una animación o un flujo cumple con los estándares de accesibilidad requeridos.

4.3. Presentación del Caso de Estudio: DialogApp

Para llevar a cabo la validación necesaria, se seleccionó como caso de estudio la evaluación de la calidad del software DialogApp. DialogApp es un prototipo funcional de aplicación móvil enfocado en el área de la salud digital (mHealth), cuyo propósito principal es derribar las barreras idiomáticas y facilitar la comunicación bidireccional entre profesionales médicos y pacientes con discapacidad auditiva.

La aplicación contiene gran parte de los elementos tecnológicos e interactivos necesarios (interfaces visuales, modelos 3D y flujos de traducción) para que los evaluadores puedan aplicar las 24 aseveraciones de la guía de uso.

Para que los evaluadores puedan evaluar el software, se les sugerirá seguir tres flujos principales de interacción que corresponden con las tres principales funcionalidades de DialogApp. Posteriormente, se procederá a evaluar el modelo y guía en base a esta experiencia. Los flujos son:

- **Traducción de texto a señas** (Text-to-Sign): Orientado al personal de salud. Permite ingresar indicaciones médicas o preguntas en formato de texto, las cuales son interpretadas y reproducidas visualmente por un avatar tridimensional en lengua de señas.
- **Traducción de señas a texto** (Sign-to-Text): Orientado al paciente sordo. Utiliza la cámara del dispositivo móvil para capturar los movimientos del usuario. El sistema es capaz de reconocer tanto señas estáticas (como letras individuales) como dinámicas (con movimiento), traduciéndolas a texto en la pantalla en tiempo real.
- **Traducción de señas a voz** (Text-to-Speech): Funciona como apoyo para el profesional médico. Una vez que la aplicación traduce las señas del paciente a texto, el sistema ofrece la opción de convertir ese texto en audio, permitiendo una recepción más natural y rápida del mensaje.

Capítulo 5

Resultados

5.1. Resultados del Instrumento (Guía de Uso)

5.1.1. Hallazgos del Caso de Estudio

La aplicación de la guía de uso por parte de los evaluadores sobre DialogApp permitió levantar un diagnóstico de esta, evidenciando su comportamiento de la aplicación frente a las cualidades propuestas de inclusión. La Tabla 5.1 expone los promedios obtenidos en cada una de las 8 cualidades evaluados por el panel de evaluadores. (Ver Anexo 4 para valores por evaluador)

A partir de los datos tabulados, el diagnóstico arroja tres escenarios claramente diferenciados:

Cualidades con buen desempeño: el instrumento logró identificar que DialogApp posee un diseño con una navegabilidad y visualmente sólido. Los atributos mejor evaluados corresponden a las interacciones básicas del sistema:

- Navegabilidad Intuitiva (4.58): El puntaje más alto de la evaluación. Los evaluadores concordaron en que el flujo principal de la aplicación es directo y no requiere esfuerzo cognitivo excesivo para que un usuario transite por sus pantallas.

Cuadro 5.1: Resultados de Evaluación Promedio por Atributo

Atributo	N° Pregunta	Promedio	Promedio del Atributo
Claridad Visual	1.1a	4.25	4.33
	1.1b	4.00	
	1.1c	4.75	
Navegabilidad Intuitiva	1.2a	4.75	4.58
	1.2b	4.25	
	1.2c	4.75	
Tolerancia a Errores	1.3a	1.50	1.92
	1.3b	1.50	
	1.3c	2.75	
Precisión Espacial	2.1a	5.00	4.50
	2.1b	4.00	
	2.1c	4.50	
Naturalidad	2.2a	2.25	1.83
	2.2b	2.00	
	2.2c	1.25	
Comprensibilidad	2.3a	3.75	3.17
	2.3b	1.75	
	2.3c	4.00	
Fiabilidad de la Información	3.1a	3.25	2.83
	3.1b	1.75	
	3.1c	3.50	
Aceptación Cultural	3.2a	4.50	3.67
	3.2b	2.75	
	3.2c	3.75	

- Precisión Espacial (4.50): Destaca particularmente la aseveración 2.1a, lo que confirma que la posición, la postura y la disposición del avatar 3D dentro de la pantalla son apropiados para la lectura de señas.
- Claridad Visual (4.33): Nos dice que el contraste y el tamaño de los elementos gráficos se ajustan a las necesidades de los usuarios.

Desempeño Moderado: existen dimensiones donde la aplicación cumple de manera general, pero presenta algunas limitaciones durante su ejecución:

- Aceptación Cultural (3.67): Aunque el sistema logra una buena representación gráfica (aseveración 3.2a con 4.50), sufre una baja en la aseveración 3.2b (2.75). Esto nos sugiere que se debe avanzar en la confianza y profesionalismo de la aplicación hacia el usuario, para que sea más aceptable para la comunidad.
- Comprensibilidad (3.17): Si bien el mensaje logra ser transmitido, se detectó una debilidad importante en la aseveración 2.3b (1.75), lo cual tiene sentido, ya que DialogApp no tiene ningún tipo de interfaz para regular la velocidad de traducción, lo que explicaría la baja puntuación en este apartado.

Atributos con bajo desempeño: la Guía de Uso expuso limitaciones en la interacción avanzada y el realismo de la traducción, áreas relevantes para la comunidad sorda:

- Naturalidad (1.83): fue el atributo peor evaluado, con una baja calificación en la aseveración 2.2c. Esto indica que aunque el avatar se posiciona bien (Precisión Espacial), sus movimientos carecen de la fluidez necesaria, resultando en transiciones robóticas que afectan la experiencia.
- Tolerancia a Errores (1.92): se asignó un bajo puntaje a la capacidad del sistema para recuperar al usuario frente a equivocaciones lo que revela una falta de retroalimentación cuando ocurre un fallo en el flujo de traducción. Esto es porque en secciones no había una forma clara de detectar un fallo, o si era detectado por la persona, no había una forma clara de recuperarse.

- Fiabilidad de la Información (2.83): el promedio de este atributo se vio severamente afectado por la baja calificación en la aseveración 3.1b (1.75), lo que nos dice que pesar de que el sistema funciona a nivel de interfaz, existen problemas en este apartado, con especial énfasis en que el sistema no entrega al usuario seguridad y/o claridad con respecto a su privacidad, independiente de que sea justificado o no.

5.1.2. Hallazgos estadísticos

Para validar que la Guía de Uso es una herramienta de medición objetiva y que sus resultados no dependen del evaluador, se midió la concordancia entre evaluadores. Previo al análisis, se definió la hipótesis nula (H_0) como la ausencia de acuerdo entre los evaluadores, frente a la hipótesis alternativa (H_1) la cual se define como la existencia de concordancia entre los evaluadores. Para esto, se aplicó el Coeficiente de Concordancia de Kendall (W) a las puntuaciones (escala de Likert de 1 a 5) asignadas de forma independiente por el panel de evaluadores ($m = 4$) a los ítems de la guía de uso ($n = 24$), donde se escribió un programa en Python que permitió realizar el cálculo (Ver Anexo 3). El análisis de datos dio como resultado un coeficiente $W = 0,752$. Para este estadístico, un valor superior a 0.60 indica un nivel de acuerdo fuerte. Además, se obtuvo un valor- $p < 0,001$, lo que nos da un nivel de confianza del cálculo mayor al 99.9 %, lo que nos sugiere que la concordancia entre los evaluadores no es producto del azar y rechaza la hipótesis nula. La correlación obtenida confirma que los 24 ítems de la pauta son claros, aplicables y reducen la subjetividad al evaluar software. Esto nos sugiere que la Guía de Uso permitió a los diversos evaluadores evaluar DialogApp y obtener resultados con una concordancia inter-evaluador fuerte.

5.2. Resultados de Validación

Una vez concluido el caso de estudio, se procedió a aplicar el segundo instrumento con el objetivo de validar el modelo y la guía de uso. Cabe destacar que, según lo declarado en el preámbulo del instrumento, los evaluadores contaban con conocimientos en metodologías de testeo, pero no así en el ámbito de accesibilidad para la comunidad sorda. (Ver Anexo 4 para

valores por evaluador)

Cuadro 5.2: Resultados de Validación Promedio por Propiedad

Propiedad	N° Pregunta	Promedio	Promedio de la Propiedad
Aplicabilidad	1	4.50	4.20
	2	4.75	
	3	3.50	
	4	3.75	
	5	4.50	
Claridad	6	4.50	4.67
	7	4.75	
	8	4.75	
Compleitud	9	4.00	4.13
	10	4.25	

- **Aplicabilidad:** La evaluación de esta propiedad presentó resultados mixtos. Por un lado, los evaluadores confirmaron que el modelo y la Guía de Uso son herramientas que permiten la evaluación o medición: destacaron que la pauta facilita la aplicación práctica (4.75), permite una evaluación integral (4.50) y requiere un nivel de esfuerzo cognitivo adecuado para un entorno de auditoría profesional (4.50). Sin embargo, las métricas orientadas a medir si el instrumento reduce la subjetividad (3.75) y facilita la identificación de barreras específicas para personas sordas (3.50) obtuvieron los puntajes más bajos. Al revisar las respuestas cualitativas de los evaluadores, se observa que la falta de conocimiento en la Lengua de Señas Chilena (LSCh) generó incertidumbre al momento de evaluar el nivel de realismo y exactitud de la traducción generada por el sistema.

- **Claridad:** Esta propiedad fue la mejor evaluada por el panel de evaluadores. Los resultados indican un consenso en que las 8 cualidades del modelo están definidos de forma lógica y sin ambigüedades (4.50). Además, los promedios más altos de esta sección confirmaron que las 24 aseveraciones se asocian correctamente a elementos visuales y de interacción directa en la interfaz (4.75), y que la escala de evaluación es intuitiva para medir el cumplimiento del software (4.75). Esto nos sugiere que el lenguaje utilizado en el instrumento es comprendido por los evaluadores.
- **Complejidad:** La evaluación de esta propiedad confirmó que el alcance del modelo es adecuado para su objetivo dado que los evaluadores validaron que los atributos propuestos cubren las dimensiones necesarias para auditar la inclusión de un producto (4.00). Además, la granularidad del modelo, es decir, su división en 24 aseveraciones, fue evaluado positivamente (4.25).

5.3. Análisis Cualitativo y Retroalimentación

Para complementar los resultados cuantitativos, el instrumento incluyó preguntas abiertas que permitieron conocer la percepción de los evaluadores sobre los desafíos que tuvieron tanto de la aplicación de la guía de uso, como de las oportunidades de mejora.

Al consultar sobre los atributos observables con mayores dificultades, hubo una tendencia a las dimensiones relacionadas directamente con la traducción: Fiabilidad de la Información, Comprensibilidad y Naturalidad. El análisis de las respuestas indica que la dificultad para evaluar estos atributos se debe a dos factores. Primero, el nivel de dominio del lenguaje: los evaluadores señalaron que, al no ser usuarios nativos de lengua de señas, carecían de los parámetros necesarios para medir la velocidad de las señas o la exactitud del vocabulario. Segundo, se identificaron factores técnicos del prototipo que dificultaron la medición. Específicamente, la falta de expresiones faciales en el avatar 3D limitó la evaluación de la naturalidad de la comunicación. A pesar de estas limitaciones, los evaluadores realizaron pruebas de validación con elementos básicos, como el abecedario dactilológico, para evaluar la fiabilidad del sistema.

En cuanto a las oportunidades para expandir y robustecer la guía de uso, las sugerencias del panel se orientaron a fortalecer la evaluación de la experiencia “complementaria”. Entre las principales propuestas destacó la necesidad de incorporar métricas que auditen el “onboarding”, es decir, la capacidad del sistema para entregar instrucciones claras sobre su propio funcionamiento antes de la interacción principal y la inclusión de parámetros de colorimetría enfocados en optimizar el contraste para aumentar el confort en el uso prolongado. Estos comentarios sugieren que futuras iteraciones podrían beneficiarse de la inclusión de material de apoyo contextual (glosarios de LSCh) y de un enfoque de evaluación interdisciplinario.

Conclusiones y Trabajo Futuro

Conclusiones Generales

El desarrollo de la presente investigación culminó en la estructuración de un modelo de calidad de software orientado a la evaluación de aplicaciones móviles para la comunidad sorda. Mediante la revisión y adaptación de estándares internacionales de accesibilidad, el trabajo logró proponer un marco teórico basado en cualidades, para luego operacionalizarlas en un instrumento (guía de uso) aplicable a entornos de evaluación de software.

Esta guía de uso permitió operacionalizar las cualidades de inclusión, transformándolos en atributos donde la idea es que sean directamente observables en la interfaz, para posibilitar la evaluación e identificación sistemática de las barreras de interacción asociadas al uso de software por parte de la comunidad sorda.

Los resultados obtenidos indican que la propuesta provee una buena base hacia la dirección de la inclusión de la comunidad sorda en el proceso de desarrollo y evaluación de la calidad de software inclusivo. Por un lado, la aplicación de la guía de uso para evaluar el producto sugiere que es actuable y apropiada para diagnosticar sistemas en fase de PMV, y que permite identificar y abordar la inclusión como un componente medible. Por otro lado, la evaluación del modelo en sí mismo nos da resultados cuantificables que nos permiten argumentar su completitud y correctitud.

Conclusiones de la Validación del Modelo y la Guía de Uso

La propuesta fue sometida a un proceso de validación estructurado en dos focos: en primer lugar, el análisis cuantitativo de la consistencia del instrumento arrojó un Coeficiente de Concordancia de Kendall de 0.752, lo que nos muestra que la guía de uso posee un alto grado de correlación entre evaluadores. Los datos obtenidos nos aportan evidencia de que la guía ayuda a disminuir la subjetividad de cada persona, permitiendo obtener cifras con una correlación significativa entre distintos evaluadores.

En segundo lugar, la evaluación del modelo y guía en sí mismos, ejecutada por el panel de evaluadores, donde sus resultados nos respaldan las propiedades de la propuesta en cuanto a su aplicabilidad, claridad y completitud, obteniendo cada propiedad un puntaje alto. Los resultados recolectados evidenciaron que el nivel de granularidad de las aseveraciones observables y la escala de calificación resultan intuitivos y operables para el perfil de evaluador. También, esta fase de validación detectó que las cualidades vinculadas directamente al conocimiento lingüístico (tales como la naturalidad de las transiciones o la exactitud del mensaje) están fuertemente condicionadas por el nivel de dominio del evaluador en lengua de señas. Esto sugiere que futuras iteraciones o la aplicación a futuro de la guía de uso requiere un enfoque interdisciplinario con personas con un conocimiento importante de lengua de señas.

Conclusiones sobre el Caso de Estudio

La aplicación práctica sobre el producto DialogApp sirvió como un caso de estudio instrumental para buscar evidencia que la guía de uso es apropiada y actuable. El análisis de los datos aportó evidencia de que el instrumento posee la granularidad necesaria para evaluar el software inclusivo orientado a una aplicación inclusiva de traducción como DialogApp.

Por una parte, la guía de uso permitió identificar y también cuantificar las fortalezas de la aplicación, las cuales se ven reflejadas en el alto cumplimiento asociado a la interacción básica, como la navegabilidad intuitiva y la precisión espacial del entorno gráfico. Por otra, el modelo pudo identificar también brechas críticas de accesibilidad, detectando limitaciones

significativas en la tolerancia a errores y en aspectos avanzados de la traducción, específicamente en la naturalidad de los movimientos del avatar 3D y la ausencia de expresiones faciales.

Según lo anterior, podemos decir que en el caso de estudio, el instrumento cumple con el propósito de entregar un diagnóstico estructurado, donde los hallazgos no se limitan a dar simplemente una aprobación o rechazo, sino que señala las áreas donde los equipos de desarrollo deben priorizar las iteraciones de diseño para alcanzar un estándar de inclusión efectivo.

Trabajo Futuro

A partir de los hallazgos obtenidos, se identifican diversas líneas de desarrollo y oportunidades de mejora para iteraciones posteriores, y buscar expandir tanto el alcance del instrumento como su aplicabilidad.

En primer lugar, en cuanto a la expansión de las aseveraciones de la Guía de Uso, la retroalimentación de los evaluadores sugiere incorporar parámetros orientados a la experiencia de usuario periférica (fuera del flujo principal), por lo que sería conveniente investigar criterios que auditen los procesos de onboarding para asegurar la correcta entrega de instrucciones previas a la interacción principal y la inclusión de directrices avanzadas de colorimetría enfocadas en evaluar el confort visual ante una exposición prolongada a la pantalla.

En segundo lugar, para que los evaluadores no dependan de su conocimiento previo en lengua de señas al momento de evaluar, se deja propuesto el diseño de un marco de evaluación interdisciplinario, que incluya personas con conocimiento aceptable de lengua de señas, donde futuras aplicaciones del modelo podrían estructurarse bajo un esquema de evaluación por pares, integrando a un ingeniero de software responsable de la auditoría estructural y a un usuario nativo o intérprete de Lengua de Señas Chilena (LSCh) encargado de evaluar la precisión semántica. Otra forma de mejorar la guía de uso sería la inclusión de un glosario técnico o material visual de apoyo que asista a los evaluadores técnicos en la medición de

atributos complejos, como la naturalidad o la fiabilidad de la traducción, y que les dé ejemplos, pero esto trae el desafío de que los evaluadores no serían capaces de probar el sistema fuera de los ejemplos sugeridos en este glosario.

Finalmente, en cuanto a la escalabilidad, se plantea la oportunidad de adaptar o evaluar este modelo de calidad más allá del contexto de las aplicaciones móviles. El marco teórico y las dimensiones estructuradas en esta investigación ofrecen una base para evaluar la inclusión en plataformas ya sean web, de escritorio u otras.

Bibliografía

- [1] M. D. D. S. Y. FAMILIA, “LEY 21303 - MODIFICA LA LEY N° 20.422, QUE ESTABLECE NORMAS SOBRE IGUALDAD DE OPORTUNIDADES E INCLUSIÓN SOCIAL DE PERSONAS CON DISCAPACIDAD, PARA PROMOVER EL USO DE LA LENGUA DE SEÑAS,” Jan. 2021.
- [2] M. J. Herrera, “Falta de inclusión, estigma social y discriminación: la compleja realidad diaria que vive una persona sorda en Chile,” Mar. 2023.
- [3] N. Alhammad, M. Alajlani, A. Abd-alrazaq, G. Epiphaniou, and T. Arvanitis, “Patients’ Perspectives on the Data Confidentiality, Privacy, and Security of mHealth Apps: Systematic Review,” *Journal of Medical Internet Research*, vol. 26, p. e50715, May 2024.
- [4] N. Wilkins, Z. Yu, N. Kelly, D. Murphy, and Y. Payano, “Comparing Sign Language AI and Human Interpreters: Evaluating Efficiency for Greater Social Sustainability,”
- [5] Y. Zhao, R. Zhang, D. Yao, and Y. Chen, “Overview of CCL24-Eval Task 10: Translation Quality Evaluation of Sign Language Avatar,”
- [6] L. C. Quandt, A. Willis, M. Schwenk, K. Weeks, and R. Ferster, “Attitudes Toward Signing Avatars Vary Depending on Hearing Status, Age of Signed Language Acquisition, and Avatar Type,” *Frontiers in Psychology*, vol. 13, p. 730917, Feb. 2022.
- [7] M. Huenerfauth, L. Zhao, E. Gu, and J. Allbeck, “Evaluation of American Sign Language Generation by Native ASL Signers,” *ACM Transactions on Accessible Computing*, vol. 1, pp. 1–27, May 2008.
- [8] S. Bennbaia, “Toward an evaluation model for signing avatars,” *Nafath*, vol. 7, May 2022.
- [9] W. Xiong, T. Yao, Q. Pan, and Z. Liu, “Usability Heuristic Evaluation for the Hearing Impaired Language Training Mobile App,” in *Design, User Experience, and Usability. Interaction Design* (A. Marcus and E. Rosenzweig, eds.), vol. 12200, pp. 294–307, Cham: Springer International Publishing, 2020. Series Title: Lecture Notes in Computer Science.

- [10] S. R. Stoyanov, L. Hides, D. J. Kavanagh, O. Zelenko, D. Tjondronegoro, and M. Mani, "Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps," *JMIR mHealth and uHealth*, vol. 3, p. e27, Mar. 2015.
- [11] ISO/IEC, "INTERNATIONAL STANDARD ISO/IEC 25010 - Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Product quality model," Nov. 2023.
- [12] World Wide Web Consortium, "Web content accessibility guidelines (WCAG) 2.2," w3c recommendation, World Wide Web Consortium, Oct. 2023. Consultado el 2025-12-15.
- [13] K. Moumane, A. Idri, F. El Aouni, J. Laghnimi, N. C. Benabdellah, and O. Hamal, "ISO/IEC 25010-based Quality Evaluation of Three Mobile Applications for Reproductive Health Services in Morocco," *Clinical and Experimental Obstetrics & Gynecology*, vol. 51, p. 88, Mar. 2024.
- [14] Y. Chinchay, J. Gomez, and G. Montoro, "Unlocking inclusive education: A quality assessment of software design in applications for children with autism," *Journal of Systems and Software*, vol. 217, p. 112164, Nov. 2024.
- [15] A. Yeratziotis and P. Zaphiris, "A Heuristic Evaluation for Deaf Web User Experience (HE4DWUX)," *International Journal of Human-Computer Interaction*, vol. 34, pp. 195–217, Mar. 2018.
- [16] C. Moreno Martínez, J. R. Piorno, J. J. Escribano Otero, and M. Guijarro Mata-García, "Responsive inclusive design (RiD): a new model for inclusive software development," *Universal Access in the Information Society*, vol. 22, pp. 893–902, Aug. 2023.
- [17] C. S. Bianchini, F. Borgia, and M. De Marsico, "A Concrete Example of Inclusive Design: Deaf-Oriented Accessibility," in *The Wiley Handbook of Human Computer Interaction* (K. L. Norman and J. Kirakowski, eds.), pp. 731–756, Wiley, 1 ed., Feb. 2018.
- [18] A. Othman, A. Dhouib, H. Chalghoumi, O. El Ghoul, and A. Al-Mutawaa, "The Acceptance of Culturally Adapted Signing Avatars Among Deaf and Hard-of-Hearing Individuals," *IEEE Access*, vol. 12, pp. 78624–78640, 2024.
- [19] F. Wasserroth, E. Avramidis, V. Czehmann, T. Kojic, F. Nunnari, and S. Möller, "Evaluation of a Sign Language Avatar on Comprehensibility, User Experience & Acceptability," in *Adjunct Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents*, (Berlin Germany), pp. 1–10, ACM, Sept. 2025.
- [20] R. L. Romero, F. Kates, M. Hart, A. Ojeda, I. Meirom, and S. Hardy, "Quality of Deaf and Hard-of-Hearing Mobile Apps: Evaluation Using the Mobile App Rating Scale (MARS) With Additional Criteria From a Content Expert," *JMIR mHealth and uHealth*, vol. 7, p. e14198, Oct. 2019.

- [21] S. Siegel and N. J. Castellan, *Nonparametric statistics for the behavioral sciences*. Boston, Mass.: McGraw-Hill, 2. ed., [reprinted] ed., 2003.
- [22] A. P. Field, “Kendall’s Coefficient of Concordance,” in *Encyclopedia of Statistics in Behavioral Science* (B. S. Everitt and D. C. Howell, eds.), Wiley, 1 ed., Apr. 2005.
- [23] P. Legendre, “Species associations: the Kendall coefficient of concordance revisited,” *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 10, pp. 226–245, June 2005.

Anexo 1: Pauta de Evaluación del Software

INSTRUCCIONES DE EVALUACIÓN

Al responder este formulario, usted acepta participar voluntariamente en esta evaluación. Los datos y opiniones recopilados serán utilizados con fines estrictamente académicos. Los resultados obtenidos podrán ser analizados y publicados. Se garantiza que el tratamiento de los datos será completamente confidencial y se mantendrá el anonimato de su identidad en toda publicación.

Lea atentamente cada afirmación y marque el círculo que mejor represente su evaluación respecto a la interfaz o avatar. Utilice la siguiente escala:

- 1 = Totalmente en desacuerdo
- 5 = Totalmente de acuerdo

Afirmación a evaluar	1	2	3	4	5
Eje 1: Usabilidad de Interfaz					
Claridad Visual					
1.1a La interfaz usa un mínimo de textos largos y utiliza los espacios en blanco adecuadamente para no saturar al usuario.	☐	☐	☐	☐	☐

Afirmación a evaluar	1	2	3	4	5
1.1b Existe un contraste adecuado entre los elementos interactivos y el fondo que facilita su visualización y lectura rápida.	☐	☐	☐	☐	☐
1.1c La pantalla está libre de elementos decorativos o animaciones que puedan generar distracción visual o sobrecarga cognitiva.	☐	☐	☐	☐	☐
Navegabilidad Intuitiva					
1.2a La aplicación utiliza iconografía estandarizada y clara que permite navegar sin depender de menús textuales complejos.	☐	☐	☐	☐	☐
1.2b Los flujos para completar tareas son directos y requieren un número mínimo de interacciones (clics o toques de pantalla).	☐	☐	☐	☐	☐
1.2c La ubicación de los botones y controles principales se mantiene consistente en todas las pantallas, reduciendo el esfuerzo de aprendizaje.	☐	☐	☐	☐	☐
Tolerancia a Errores					
1.3a El sistema emite respuestas visuales o de vibración (hápticas) claras e inmediatas cuando el usuario realiza una acción.	☐	☐	☐	☐	☐
1.3b Ante un error, el sistema guía al usuario de manera visual, evitando entregar únicamente mensajes de fallo en formato de texto.	☐	☐	☐	☐	☐

Afirmación a evaluar	1	2	3	4	5
1.3c La interfaz ofrece la opción de cancelar o deshacer fácilmente una acción crítica antes de que se ejecute de manera definitiva.	☐	☐	☐	☐	☐
Eje 2: Calidad del mensaje					
Precisión Espacial					
2.1a La configuración de las manos y la orientación de las palmas son gramaticalmente correctas respecto a la seña emitida.	☐	☐	☐	☐	☐
2.1b El emisor del mensaje utiliza correctamente el espacio tridimensional para situar sujetos y acciones dentro de la oración.	☐	☐	☐	☐	☐
2.1c Las trayectorias de movimiento entre un punto espacial y otro son exactas y no generan deformaciones anatómicas en la animación.	☐	☐	☐	☐	☐
Naturalidad					
2.2a Los movimientos corporales y la transición entre las distintas señas se perciben fluidos y evitan la rigidez robótica.	☐	☐	☐	☐	☐
2.2b Se sincronizan adecuadamente las expresiones faciales y corporales con el significado de la seña.	☐	☐	☐	☐	☐
2.2c El ritmo de la traducción y las pausas entre oraciones imitan la cadencia de una conversación humana real.	☐	☐	☐	☐	☐
Comprensibilidad					

Afirmación a evaluar	1	2	3	4	5
2.3a El sistema permite diferenciar con claridad visual aquellas señas que tienen movimientos o configuraciones muy similares.	☐	☐	☐	☐	☐
2.3b La interfaz ofrece herramientas para repetir el mensaje o ajustar su velocidad sin que se pierda la legibilidad de la seña.	☐	☐	☐	☐	☐
2.3c El mensaje conserva su intención original sin perder significado durante la traducción.	☐	☐	☐	☐	☐
Eje 3: Valor y Aceptación					
Fiabilidad de la Información					
3.1a El sistema traduce con precisión y sin ambigüedades la terminología técnica o crítica propia del contexto de la aplicación.	☐	☐	☐	☐	☐
3.1b En los flujos que gestionan datos sensibles, se transmite al usuario seguridad y claridad respecto a la protección de su privacidad.	☐	☐	☐	☐	☐
3.1c La herramienta es consistente y no entrega traducciones contradictorias al interpretar un mismo concepto en distintas pantallas.	☐	☐	☐	☐	☐
Aceptación Cultural					
3.2a La traducción respeta fielmente la gramática y el vocabulario local, sin mezclar modismos extranjeros.	☐	☐	☐	☐	☐

Afirmación a evaluar	1	2	3	4	5
3.2b El diseño estético general transmite el profesionalismo adecuado para el dominio de la aplicación, generando confianza en el usuario.	☐	☐	☐	☐	☐
3.2c La apariencia del avatar resulta empática y evita el uso de estereotipos o elementos visuales que puedan resultar extraños para la comunidad local.	☐	☐	☐	☐	☐

Anexo 2: Pauta de Evaluación del Modelo

INSTRUCCIONES DE EVALUACIÓN

Al responder este formulario, usted acepta participar voluntariamente en esta evaluación. Los datos y opiniones recopilados serán utilizados con fines estrictamente académicos. Los resultados obtenidos podrán ser analizados y publicados. Se garantiza que el tratamiento de los datos será completamente confidencial y se mantendrá el anonimato de su identidad en toda publicación.

Lea atentamente cada afirmación y marque el círculo que mejor represente su evaluación. Para las preguntas finales, escriba su respuesta en el espacio asignado. Utilice la siguiente escala para las afirmaciones:

- 1 = Totalmente en desacuerdo
- 5 = Totalmente de acuerdo

1. Profesión o actividad actual (¿A qué se dedica?):

(Indique su cargo, rol o área de desempeño)

2. Experiencia técnica y en inclusión:

(Describa brevemente su experiencia en desarrollo, QA o evaluación de software en general, y comente si ha tenido experiencia previa trabajando o interactuando con software inclusivo o de accesibilidad).

Afirmación / Pregunta a Evaluar	1	2	3	4	5
Aplicabilidad					
1. El modelo de calidad propuesto me permitiría ejecutar en la práctica una evaluación integral sobre un producto de software inclusivo.	☐	☐	☐	☐	☐
2. La guía de uso (pauta de 24 aseveraciones) resulta ser una herramienta práctica que facilita la aplicación de este modelo durante la auditoría técnica del software.	☐	☐	☐	☐	☐
3. El uso de este modelo facilita la identificación de barreras de accesibilidad específicas para personas sordas, más allá de la usabilidad general.	☐	☐	☐	☐	☐
4. El modelo reduce la subjetividad del evaluador al momento de determinar si una aplicación cumple con los estándares de inclusión.	☐	☐	☐	☐	☐
5. El tiempo y esfuerzo cognitivo requerido para aplicar este modelo es razonable para una evaluación técnica de software.	☐	☐	☐	☐	☐
Claridad					
6. Los 8 atributos que componen el modelo están definidos de manera clara, lógica y sin ambigüedades.	☐	☐	☐	☐	☐
7. Las aseveraciones utilizadas para medir cada atributo se basan en elementos observables en la interfaz, facilitando su evaluación técnica.	☐	☐	☐	☐	☐

Afirmación / Pregunta a Evaluar (Continuación)	1	2	3	4	5
8. La escala de evaluación propuesta resultó intuitiva y adecuada para calificar el nivel de cumplimiento de cada atributo en el software.	☐	☐	☐	☐	☐
Compleitud					
9. Los atributos (qualities) del modelo son apropiados y abarcan las dimensiones fundamentales para determinar si un software es verdaderamente inclusivo para la comunidad sorda.	☐	☐	☐	☐	☐
10. El nivel de granularidad (desglose en atributos) del modelo es el adecuado para no pasar por alto fallos críticos de diseño.	☐	☐	☐	☐	☐
Retroalimentación					
11. ¿Cuál de los 8 atributos le resultó más difícil de interpretar o medir, y por qué?					
12. ¿Qué atributo, métrica o aspecto adicional incorporaría en una futura versión de este modelo de calidad?					

Anexo 3: Coeficiente de Kendall

Coeficiente de Kendall (W)

El coeficiente W de Kendall es una medida estadística que aplicada al caso de estudio permite determinar el grado de acuerdo o consenso entre varios individuos que califican un mismo conjunto de elementos.[21] A diferencia de otras métricas, no asume que los datos tengan una distribución normal, por lo que es aplicable para analizar resultados obtenidos a través de escalas ordinales como Likert (evaluaciones del 1 al 5). El valor de W fluctúa entre 0 y 1. Un valor cercano a 0 indica que no existe acuerdo entre los evaluadores (cada uno evaluó con un criterio completamente distinto), mientras que un valor 1 representa una concordancia perfecta, donde todos los evaluadores ordenaron sus resultados exactamente de la misma manera. [22]

El coeficiente se calcula de la siguiente manera: $W = \frac{12S}{m^2(n^3-n)-mT}$

Donde aplicándolo al caso de estudio, m es la cantidad de evaluadores, n el número de aseveraciones evaluadas, SS es la suma de las desviaciones al cuadrado de los rangos respecto a su promedio (variabilidad) y T es el factor de corrección por empates (se calcula como la suma de $(t^3 - t)$ por cada grupo de notas repetidas. [22]

Valor P (p-value)

El valor p es una probabilidad condicional que funciona como un control de calidad para validar el resultado de W, y representa la probabilidad de obtener un coeficiente de acuerdo igual o mayor al calculado, si se asume la hipótesis nula. Dado que el instrumento cuenta con más de 7 preguntas ($n > 7$), el cálculo del valor p no se realiza de forma directa, sino mediante una aproximación a la distribución Chi-cuadrado (χ^2)[23]. La transformación del coeficiente de Kendall para esta distribución se rige por la fórmula: $\chi^2 = m * (n - 1) * W$

Código para el cálculo de W

Para el cálculo del coeficiente se usó un código programado en Python junto a las librerías Numpy, Scipy y Pandas. El código usado para el cálculo es el siguiente:

```
import pandas as pd
import numpy as np
import scipy.stats as stats
# 1. Carga de datos desde excel
archivo_excel = 'datos_instrumento1.xlsx'
df = pd.read_excel(archivo_excel)
# Limpieza de datos
df = df.dropna()
datos = df.select_dtypes(include=[np.number]).values
# n = cantidad de aseveraciones (filas), m = cantidad de evaluadores (columnas)
n, m = datos.shape
# 2. Obtener rangos y manejo de empates por cada evaluador
rangos = np.array([stats.rankdata(datos[:, j]) for j in range(m)]).T
# 3. Suma de rangos por cada ítem
R_i = rangos.sum(axis=1)
R_media = m * (n + 1) / 2.0
S = np.sum((R_i - R_media) ** 2)
# 4. Penalización por empates
T = 0
for j in range(m):
    _, counts = np.unique(datos[:, j], return_counts=True)
    T += np.sum(counts**3 - counts)
# 5. Fórmula del coeficiente W
denominador = (m**2 * (n**3 - n)) - (m * T)
W = (12 * S) / denominador if denominador != 0 else 0
# 6. Aproximación a Chi-cuadrado y cálculo de la integral para el p-valor
chi_cuadrado = m * (n - 1) * W
p_valor = stats.chi2.sf(chi_cuadrado, n - 1)
# 7. Salida
print(f"Total de evaluadores (m) : {m}")
print(f"Total de aseveraciones (n): {n}")
print(f"W de Kendall calculado : {W:.3f}")
print(f"Valor P (p-value) : {p_valor:.4f}")
```

Anexo 4: Valores obtenidos en la sesión de Evaluación

Evaluación de la Calidad del Software

Cuadro A4.1: Valores obtenidos de la Evaluación de la Calidad del Software

Cualidad	Pregunta	Evaluable 1	Evaluable 2	Evaluable 3	Evaluable 4
Claridad Visual	1.1a	5	3	5	4
	1.1b	5	3	4	4
	1.1c	5	5	4	5
Navegabilidad Intuitiva	1.2a	5	4	5	5
	1.2b	3	5	5	4
	1.2c	5	5	4	5
Tolerancia a Errores	1.3a	1	2	2	1
	1.3b	1	3	1	1
	1.3c	4	4	1	2
Precisión Espacial	2.1a	5	5	5	5

Cualidad	Pregunta	Evaluador 1	Evaluador 2	Evaluador 3	Evaluador 4
	2.1b	4	4	4	4
	2.1c	4	5	5	4
Naturalidad	2.2a	2	2	3	2
	2.2b	3	2	1	2
	2.2c	2	1	1	1
Comprensibilidad	2.3a	5	3	3	4
	2.3b	1	1	4	1
	2.3c	5	3	4	4
Fiabilidad de la Información	3.1a	4	2	3	4
	3.1b	2	2	1	2
	3.1c	4	3	3	4
Aceptación Cultural	3.2a	5	4	4	5
	3.2b	3	2	3	3
	3.2c	4	2	5	4

Evaluación del Modelo y Guía de Uso

Cuadro A4.2: Valores obtenidos de la Evaluación del Modelo y Guía de Uso

Aseveración	Evaluador 1	Evaluador 2	Evaluador 3	Evaluador 4
El modelo de calidad propuesto me permitiría ejecutar en la práctica una evaluación integral sobre un producto de software inclusivo.	5	3	5	5
La guía de uso (pauta de 24 aseveraciones) resulta ser una herramienta práctica que facilita la aplicación de este modelo durante la auditoría técnica del software.	5	4	5	5
El uso de este modelo facilita la identificación de barreras de accesibilidad específicas para personas sordas, más allá de la usabilidad general.	5	2	4	3
El modelo reduce la subjetividad del evaluador al momento de determinar si una aplicación cumple con los estándares de inclusión.	4	3	4	4
El tiempo y esfuerzo cognitivo requerido para aplicar este modelo es razonable para una evaluación técnica de software.	5	4	5	4

Aseveración	Evaluador 1	Evaluador 2	Evaluador 3	Evaluador 4
Los 8 atributos que componen el modelo están definidos de manera clara, lógica y sin ambigüedades.	5	5	4	4
Las aseveraciones utilizadas para medir cada atributo se basan en elementos observables en la interfaz, facilitando su evaluación técnica.	4	5	5	5
La escala de evaluación propuesta resultó intuitiva y adecuada para calificar el nivel de cumplimiento de cada atributo en el software.	5	4	5	5
Los atributos (qualities) del modelo son apropiados y abarcan las dimensiones fundamentales para determinar si un software es verdaderamente inclusivo para la comunidad sorda.	5	2	5	4
El nivel de granularidad (desglose en atributos) del modelo es el adecuado para no pasar por alto fallos críticos de diseño.	5	3	5	4