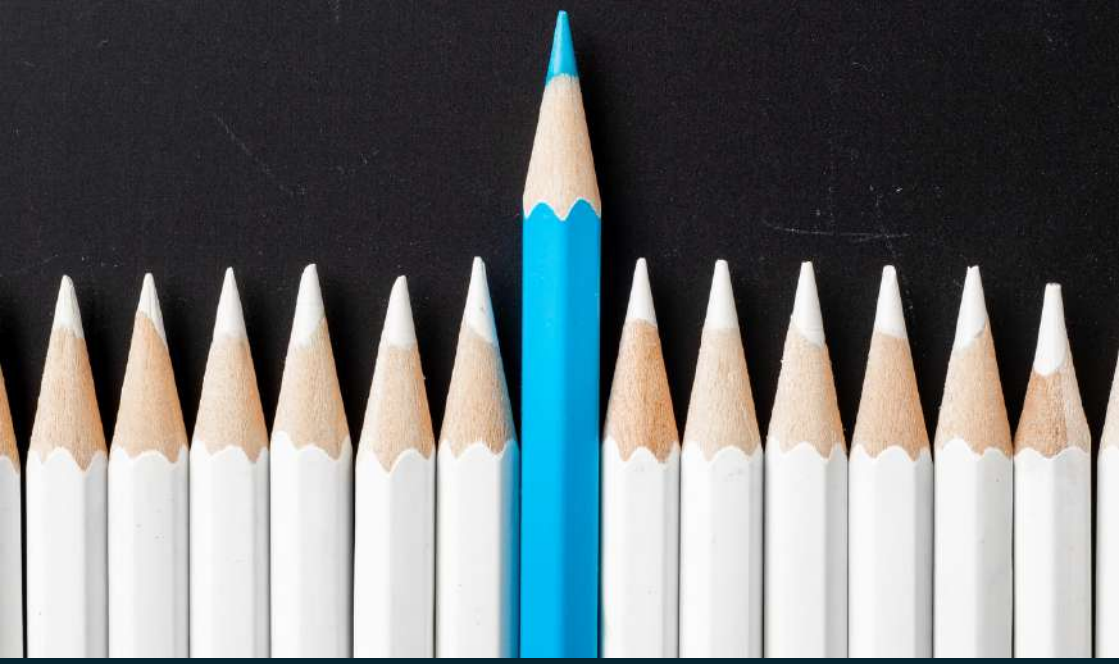
ARTI EDITORIAL

DISEÑO EFICIENTE DE **ENCUESTAS DIGITALES** EN ESTUDIOS **ADMINISTRATIVOS:**

EXTENSIÓN, TIEMPO DE RESPUESTA Y CALIDAD DE LOS DATOS



JORGE ANDRÉS IZAGUIRRE OLMEDO

DISEÑO EFICIENTE DE ENCUESTAS DIGITALES EN ESTUDIOS ADMINISTRATIVOS: *EXTENSIÓN, TIEMPO DE RESPUESTA Y CALIDAD DE LOS DATOS*

Jorge Andrés Izaguirre Olmedo
Universidad Internacional del Ecuador - UIDE

Información de la obra
Publicación digital
Primera edición – Septiembre del 2026
ISBN: 978-9907-9600-0-6
DOI: http://doi.org/10.5281/zenodo.22945727
Materia: 001.4 - Investigación
Colección: Investigación
No. pág.: 100

Cita sugerida: Izaguirre, J. (2026). Diseño eficiente de encuestas digitales en estudios administrativos: Extensión, tiempo de respuesta y calidad de los datos. *ARTI EDITORIAL*. http://doi.org/10.5281/zenodo.22945727

Equipo editorial

Mgtr. Roberto Romero Chévez
Director General

Mgtr. Víctor Sancán Chávez
Editor en Jefe

Téc. Winston Morán Párraga
Edición y Maquetación

Información Editorial
Arti Editorial
www.artieditorial.org
e-mail: info@artieditorial.org
Guayaquil – Ecuador

ÍNDICE GENERAL

ÍNDICE DE TABLAS

DEDICATORIA

A mi abuela Edith, cuya memoria permanece viva en mi corazón y cuyo ejemplo, amor y enseñanzas continúan acompañando cada paso de mi vida. Aunque ya no esté físicamente, su recuerdo es una presencia permanente y una fuente de inspiración.

A mi mujer, Lissette, por caminar a mi lado, por su amor, comprensión y apoyo incondicional, especialmente en aquellos momentos en los que alcanzar esta meta implicó esfuerzo, sacrificio y perseverancia.

Y a mi hijo Alejandro, por ser una de mis mayores motivaciones y por recordarme, con su existencia, que todo esfuerzo adquiere un sentido más profundo cuando se hace pensando en quienes amamos.

AGRADECIMIENTO

Expreso mi más sincero agradecimiento a mi familia, por su apoyo incondicional, paciencia y constante motivación durante el proceso de construcción de esta obra; a mis profesores del Doctorado, por sus enseñanzas, orientación y exigencia académica, que contribuyeron significativamente a mi formación y a la consolidación de este trabajo; y a mis compañeros doctorandos, por las experiencias compartidas, las discusiones académicas, el intercambio de ideas y el acompañamiento a lo largo de este camino. A todos ellos, mi reconocimiento y gratitud por haber sido parte fundamental de este proceso y por contribuir, desde sus diferentes espacios, a que este libro pudiera convertirse en una realidad.

PRÓLOGO

La calidad de una encuesta administrativa comienza antes de abrir el formulario. Se construye al decidir qué información se necesita, quién puede proporcionarla y cuánto esfuerzo resulta razonable solicitar. Una organización puede reunir cientos de respuestas y, aun así, desconocer si sus preguntas fueron comprendidas, si representan adecuadamente las variables estudiadas o si quienes respondieron tuvieron condiciones para hacerlo con atención. La disponibilidad de plataformas digitales facilita la recolección, pero mantiene vigente la responsabilidad de diseñar mediciones interpretables.

Este libro examina la relación entre extensión del cuestionario, tiempo de respuesta y calidad de los datos en estudios administrativos. Los fundamentos teóricos se articulan con una investigación cuantitativa realizada con trabajadores de los sectores público y privado de Ecuador. La eficiencia se entiende como la obtención de información útil con una carga proporcionada para el participante, conservando la cobertura conceptual necesaria para responder la pregunta de investigación.

Una conclusión orienta el desarrollo: la extensión debe justificarse junto con el contenido, el tiempo real de respuesta y los indicadores de calidad. Las comparaciones entre 30, 40, 50 y 65 preguntas ofrecen una referencia localizada para el diseño, pero no demuestran que exista una cantidad óptima aplicable a toda población o investigación. La decisión exige atender al tipo de ítems, las condiciones de aplicación, la experiencia del participante y la validez de las puntuaciones obtenidas.

Los capítulos iniciales presentan los fundamentos teóricos. El tercero desarrolla la metodología cuantitativa; el cuarto expone los resultados de las versiones aplicadas; y el quinto

reúne la discusión, las limitaciones, las conclusiones y las recomendaciones de la investigación.

INTRODUCCIÓN

Las investigaciones administrativas utilizan encuestas digitales para recoger información sobre experiencias y procesos organizacionales. La facilidad de distribuir un formulario y añadir preguntas amplía las posibilidades de investigación, pero también plantea una exigencia metodológica: conocer si las respuestas mantienen condiciones suficientes de atención, consistencia e integridad a medida que aumenta la extensión del instrumento.

Este libro aborda ese problema mediante fundamentos teóricos y una investigación cuantitativa centrada en la extensión del cuestionario y la calidad de respuesta. El estudio empleó cuatro versiones adaptadas del Multifactor Leadership Questionnaire, con 30, 40, 50 y 65 preguntas. Participaron 473 trabajadores de los sectores público y privado de Ecuador, reclutados mediante conveniencia y bola de nieve. El trabajo de campo se realizó entre noviembre de 2024 y diciembre de 2025.

La comparación examinó un índice compuesto por siete señales del proceso de respuesta, además del tiempo de llenado, la carga percibida y características personales. El contenido de liderazgo proporcionó una base común para los instrumentos; la pregunta central fue metodológica. Se buscó conocer qué diferencias de calidad se observaban entre las versiones y cómo se relacionaba el índice con las variables analizadas dentro de cada grupo.

Los resultados mostraron mayores medias de calidad en las versiones de 30 y 40 preguntas que en las de 50 y 65, conforme a los contrastes realizados. También identificaron relaciones con el tiempo, la carga percibida, la edad y el género. Su interpretación considera el muestreo voluntario, los distintos

momentos de aplicación, las reglas del índice y las condiciones de los análisis estadísticos.

Los capítulos iniciales explican los conceptos necesarios para comprender las encuestas digitales y los indicadores de respuesta. La exposición continúa con la metodología efectivamente aplicada, la presentación de resultados y su discusión con la literatura. El cierre formula conclusiones y recomendaciones vinculadas con la investigación, sin convertir las diferencias observadas en límites universales de longitud o duración.

La obra se dirige a investigadores, docentes, estudiantes y profesionales que utilizan encuestas en ciencias administrativas. Su propósito es contribuir a decisiones de medición que relacionen la cantidad de información solicitada con la calidad de los datos obtenidos y con la experiencia del participante. El estudio ofrece una referencia empírica situada para evaluar esos aspectos en nuevas aplicaciones.

CAPÍTULO I

FUNDAMENTOS DE LAS ENCUESTAS DIGITALES EN LA INVESTIGACIÓN ADMINISTRATIVA

Una encuesta administrativa se justifica por el conocimiento que permite obtener sobre una organización y por la calidad de las decisiones que ese conocimiento puede sostener. Su diseño exige relacionar una pregunta de investigación con una población, una forma de observación y un plan de análisis. La plataforma digital interviene en esa relación: determina cómo se muestran los reactivos, qué recorridos siguen los participantes y qué información del proceso queda registrada. Estas posibilidades deben integrarse al diseño desde el comienzo.

El capítulo sitúa la encuesta dentro del trabajo de investigación y examina las condiciones cognitivas y organizacionales de la respuesta. La exposición presta atención al vínculo entre el concepto que se desea medir y la experiencia concreta de la persona que contesta. Este vínculo explica por qué un cuestionario breve puede producir información insuficiente y por qué uno conceptualmente amplio puede fracasar si impone una carga difícil de sostener.

1.1. La encuesta como estrategia de conocimiento organizacional

La investigación administrativa estudia procesos, decisiones y relaciones que operan en diferentes niveles. Para convertirlos en información analizable, debe establecer qué aspectos requieren autoinforme y cuáles pueden observarse mediante registros, entrevistas u otras fuentes. La encuesta ocupa un lugar específico dentro de esa combinación.

1.1.1. De la pregunta de investigación a la decisión administrativa

El primer paso consiste en formular una pregunta cuyo alcance pueda reconocerse en las respuestas. Preguntar por la satisfacción del personal requiere precisar si interesa la valoración general del trabajo, la experiencia con una jefatura o un proceso específico. Estas formulaciones remiten a contenidos distintos. Un formulario que reúne opiniones generales y luego las interpreta como medidas de desempeño introduce una distancia difícil de corregir mediante técnicas estadísticas.

La decisión administrativa asociada ayuda a delimitar el contenido. Si una institución necesita revisar su procedimiento de solicitudes internas, conviene conocer qué información reciben los usuarios, cómo identifican al responsable y qué sucede cuando aparece una demora. Una pregunta sobre la imagen global de la institución puede aportar contexto, pero no sustituye esas observaciones. El valor de cada reactivo depende de la relación entre su respuesta y la inferencia que se pretende realizar.

También debe distinguirse el estudio exploratorio de la evaluación de una intervención. El primero puede identificar experiencias que requieren investigación adicional. La segunda necesita criterios previos, observaciones comparables y un diseño que permita valorar cambios. Aplicar una encuesta después de modificar un proceso no demuestra por sí mismo que la modificación haya producido las valoraciones observadas. El momento de aplicación, la composición de la muestra y otros cambios organizacionales pueden intervenir.

En una situación didáctica, un departamento recibe quejas sobre retrasos. Un instrumento centrado únicamente en

satisfacción global arroja una puntuación intermedia, pero no permite saber si el problema está en las instrucciones iniciales, la asignación de responsables o la comunicación del resultado. Otro instrumento, organizado por etapas del trámite, permite localizar esas experiencias. La segunda opción ofrece mayor utilidad para revisar el proceso, aunque su número de preguntas sea semejante al de la primera.

La calidad conceptual exige documentar qué significa la puntuación y con qué evidencias se relaciona. Flake et al. examinan la necesidad de sostener la interpretación de las mediciones con procedimientos explícitos de validación de constructo. Esta exigencia resulta especialmente relevante cuando una denominación familiar, como compromiso o eficiencia, se utiliza sin aclarar qué contenido concreto representan los ítems (Flake et al., 2017).

El producto inicial puede ser una ficha breve que contenga la pregunta de investigación, la decisión que se informará, la población, el período de referencia y la unidad de análisis. Esta ficha no es un trámite adicional. Permite detectar solicitudes de preguntas que no corresponden al objetivo y defender una extensión razonable frente a la acumulación de intereses. Si un reactivo no tiene una función analítica identificable, su inclusión requiere una justificación específica.

1.1.2. Diferencias entre encuesta, cuestionario e instrumento

La encuesta comprende el proceso de producción de información: selección y contacto de participantes, administración del cuestionario, seguimiento, codificación y análisis. El cuestionario es el conjunto organizado de preguntas

e instrucciones que forma parte de ese proceso. El instrumento de medición incluye además las reglas que vinculan las respuestas con indicadores o puntuaciones. Utilizar estos términos con precisión permite reconocer que un formulario técnicamente funcional todavía puede carecer de un modelo de medición adecuado.

Una escala es una forma particular de instrumento que utiliza varios reactivos para representar un atributo. Sus reglas de puntuación deben corresponder a la estructura del constructo. Una lista de preguntas sobre distintos servicios no constituye automáticamente una escala unidimensional. Si se suman respuestas sobre asuntos diferentes, el resultado puede ser un índice compuesto, cuya interpretación exige explicar la selección de componentes y el criterio de agregación.

Esta distinción importa al evaluar la extensión. En una escala, retirar preguntas puede reducir la cobertura del atributo o alterar su estructura. En un cuestionario que integra varios módulos, una reducción puede consistir en retirar un módulo no esencial, preservar otro completo o distribuir contenidos entre muestras. El criterio de eficiencia se aplica a la función de los componentes; contar preguntas sin identificar su función ofrece una descripción incompleta del esfuerzo solicitado.

El desarrollo de cuestionarios requiere decisiones conceptuales, redacción, evaluación y revisión iterativa. Artino et al. presentan una secuencia de elaboración que destaca la revisión de literatura, la construcción de preguntas y el examen del proceso de respuesta. Aunque su trabajo se sitúa en educación médica, el principio metodológico puede trasladarse a la investigación administrativa: cada pregunta debe probarse con quienes interpretarán el formulario. La aplicación sectorial

que se desarrolla aquí corresponde a una adaptación didáctica de ese principio (Artino et al., 2014).

Un estudio sobre atención al usuario puede emplear un cuestionario con datos de contexto, una escala de claridad de información y una pregunta abierta sobre dificultades. La escala genera una puntuación específica; la pregunta abierta requiere categorías de análisis; los datos de contexto permiten describir a quienes respondieron. No resulta adecuado reunir todos esos elementos en una única alfa de Cronbach o atribuirles un significado común porque aparezcan en la misma pantalla.

La documentación debería distinguir la versión del cuestionario y la versión de cada escala incorporada. Una modificación de colores puede afectar la experiencia de uso sin cambiar el contenido conceptual; una modificación del período de referencia altera la pregunta. Ambas deben registrarse, pero sus consecuencias son diferentes. Esta separación facilita identificar qué se ha evaluado cuando se afirma que un instrumento funciona de manera aceptable.

1.1.3. Población, unidad de análisis e informante

La población reúne las unidades sobre las que interesa producir conocimiento. La unidad de análisis indica a qué entidad se atribuye una conclusión: una persona, un equipo, una organización o un proceso. El informante es quien aporta los datos. Estas tres definiciones pueden coincidir, pero frecuentemente difieren. Una persona responsable de un departamento puede informar sobre su unidad, mientras que un trabajador puede describir su experiencia individual dentro de ella.

La diferencia exige cuidar el lenguaje de los reactivos. La afirmación «en mi trabajo recibo instrucciones claras» recoge una experiencia personal. La afirmación «el departamento dispone de procedimientos documentados» solicita conocimiento sobre un atributo colectivo. La segunda puede requerir un informante con acceso a documentación. Si cualquier integrante responde por intuición, la variación de las respuestas puede reflejar diferencias de información y no diferencias reales entre departamentos.

Tampoco debe interpretarse automáticamente una media de respuestas individuales como característica compartida del grupo. Para hablar de clima o de prácticas colectivas es necesario justificar la agregación, considerar cuántas personas respondieron en cada unidad y examinar la variación entre sus percepciones. Una media departamental basada en pocos participantes puede ocultar desacuerdos importantes. El cuestionario y el plan de análisis deben anticipar estas necesidades.

En encuestas digitales, el acceso al enlace no garantiza pertenencia a la población. La difusión abierta facilita incorporar personas fuera del ámbito previsto y dificulta conocer cuántas unidades elegibles recibieron una invitación. Un filtro inicial puede preguntar por la experiencia necesaria para responder, evitando pedir datos personales que no aporten a esa verificación. Cuando el marco muestral existe, la selección y el seguimiento deben aprovecharlo de manera proporcionada.

Cornesse y Blom comparan la calidad de respuesta en paneles probabilísticos y no probabilísticos. Su trabajo recuerda que el mecanismo de selección y el comportamiento de respuesta son dimensiones que deben examinarse por

separado. Una base puede contener respuestas cuidadosas y, al mismo tiempo, presentar problemas de representatividad. La evaluación de la atención individual no resuelve la forma en que las personas llegaron a participar (Cornesse y Blom, 2023).

La ficha de población debe indicar criterios de inclusión, cobertura geográfica, condición laboral y período de experiencia. Si el interés se centra en trabajadores que tramitaron solicitudes durante el último mes, quienes no realizaron ese proceso necesitan un recorrido distinto. Permitir la opción «no tuve esa experiencia» evita obligar a emitir una valoración. Esta decisión protege la interpretación del indicador y reduce trabajo innecesario para el participante.

1.1.4. Variables administrativas y alcance del autoinforme

Las variables administrativas abarcan experiencias, actitudes, comportamientos y resultados. Cada tipo requiere una estrategia de observación diferente. La percepción de apoyo puede estudiarse mediante autoinforme; el tiempo registrado de resolución de una solicitud puede obtenerse del sistema; una práctica de coordinación puede contrastarse con actas o registros de seguimiento. Seleccionar la encuesta por conveniencia técnica no elimina la necesidad de valorar si constituye la fuente apropiada.

En estudios de liderazgo, el informante puede evaluar a su superior, describir su propio comportamiento o informar sobre experiencias compartidas. Estas perspectivas no son intercambiables. La autoevaluación incorpora acceso privilegiado a intenciones y dificultades, pero también puede estar influida por la presentación personal. La evaluación de

una jefatura describe conductas observadas desde una relación específica. El significado de la puntuación depende de esa posición y del período considerado.

El manuscrito que fundamenta este libro utiliza adaptaciones de preguntas de liderazgo como contexto para estudiar la calidad de respuesta. Esta elección no convierte sus resultados en una validación completa de las versiones reducidas como medidas de estilos de liderazgo. El interés metodológico se dirige al comportamiento de respuesta frente a diferentes extensiones. Mantener esta distinción impide reutilizar esas versiones como instrumentos sustantivos equivalentes sin reunir evidencia adicional.

La fuente común de los datos puede introducir relaciones que reflejen parcialmente el procedimiento de medición. Si una misma persona valora al mismo tiempo la calidad de la comunicación y el desempeño del equipo con una redacción semejante, ambas puntuaciones pueden compartir influencias del contexto de respuesta. Podsakoff et al. revisan este problema y sus posibles remedios. La prevención comienza con el diseño y con la elección de fuentes, aunque ninguna solución aislada elimina todas las formas de sesgo (Podsakoff et al., 2003).

Una encuesta sobre prácticas de gestión puede ganar precisión al preguntar por experiencias observables. «Durante las últimas cuatro semanas, las tareas que recibí indicaban una fecha de entrega» define un contenido más delimitado que «la gestión es eficiente». La segunda formulación permite distintas interpretaciones y mezcla criterios de valoración. La primera todavía requiere revisión cognitiva, pero hace visible qué situación debe recordar el participante.

Los resultados deben denominarse según lo que se observó. Si se recopilan percepciones sobre claridad, el informe hablará de claridad percibida. Si se compara esa medida con registros de errores, podrá analizarse una relación entre ambas fuentes. Atribuir eficacia real a una percepción exige un argumento adicional. Esta disciplina terminológica permite conectar los cuestionarios con la gestión sin prometer más de lo que sus datos pueden sostener.

1.1.5. Calidad de los datos y error total de encuesta

La calidad de los datos se refiere a su aptitud para un uso definido. En una encuesta, depende de la cobertura de la población, la selección, la participación, la medición y el procesamiento. El enfoque de error total permite estudiar estos componentes de forma conjunta. Groves y Lyberg muestran la evolución de esta perspectiva y su utilidad para comprender que una buena estimación puede verse afectada por errores diferentes a la variación muestral (Groves y Lyberg, 2010).

Un error de cobertura aparece cuando partes de la población quedan fuera del procedimiento de acceso. El error de no respuesta depende de cómo difieren quienes participan y quienes no lo hacen respecto de los temas estudiados. El error de medición surge cuando las respuestas se apartan del contenido que se pretende captar. El procesamiento agrega otras posibilidades, como códigos invertidos, duplicaciones o uniones incorrectas de archivos.

Estos problemas requieren evidencias diferentes. Una tasa de finalización alta informa que muchos de quienes comenzaron llegaron al cierre, pero no demuestra que la población esté representada. Un indicador de consistencia

interna describe una propiedad de un conjunto de respuestas, pero no acredita ausencia de cobertura desigual. Del mismo modo, eliminar registros sospechosos no corrige un cuestionario que omitió una dimensión sustantiva del constructo.

Tabla 1.
Componentes de calidad y evidencias de revisión

Componente	Pregunta de revisión	Evidencia útil
Cobertura	¿Quién puede acceder a la invitación?	Marco de contacto y canales utilizados
Participación	¿Quiénes comienzan y quiénes terminan?	Estados finales y denominadores explícitos
Medición	¿Qué interpretan las personas al responder?	Entrevistas cognitivas y estructura de las puntuaciones
Procesamiento	¿Cómo se transformaron las respuestas?	Diccionario de datos y registro de cambios
Uso	¿Qué conclusión admite la información?	Alcance de la muestra y evidencia de validez

Fuente: Elaboración propia.

La evaluación de eficiencia debe considerar posibles compensaciones. Exigir respuesta en todos los campos puede disminuir las celdas vacías entre cuestionarios enviados, pero también aumentar respuestas inventadas o abandonos. Reducir la longitud puede facilitar la participación y, simultáneamente, dejar una dimensión representada por muy pocos reactivos. Examinar un único indicador conduce a decisiones incompletas cuando las consecuencias recaen en otros componentes de calidad.

Una práctica útil consiste en declarar, antes de la recolección, qué riesgos resultan más relevantes para la pregunta del estudio y cómo se observarán. La investigación sobre calidad de respuesta necesita conservar indicadores del proceso, incluidos registros incompletos cuando estén disponibles y sea procedente utilizarlos. En cambio, una investigación sobre satisfacción puede requerir un conjunto analítico distinto. La selección depende del propósito y debe poder explicarse sin alterar retrospectivamente los criterios para favorecer un resultado.

1.2. Procesos de respuesta y condiciones del entorno digital

El dato de una encuesta es el resultado de una interacción. La persona interpreta una pregunta, busca información, forma un juicio y selecciona una opción en una interfaz concreta. El diseño eficiente facilita ese recorrido y observa las condiciones que pueden interrumpirlo.

1.2.1. Comprensión, recuperación, juicio y selección de respuestas

Comprender una pregunta implica reconocer sus términos, su referencia temporal y la situación a la que se aplica. La expresión «recibo apoyo suficiente» puede referirse a recursos materiales, asesoramiento o respaldo ante conflictos. Si el instrumento no delimita esa referencia, dos participantes pueden elegir el mismo valor por razones diferentes. La consistencia aparente de las puntuaciones no elimina esta ambigüedad semántica.

La recuperación consiste en traer a la memoria experiencias pertinentes. Una pregunta sobre «el último trámite realizado» orienta hacia un episodio identificable. Una pregunta sobre «lo que normalmente ocurre» exige una síntesis de experiencias y puede quedar dominada por acontecimientos recientes o particularmente intensos. El período de referencia debe corresponder a la frecuencia del fenómeno y a las posibilidades razonables de recuerdo del informante.

En el juicio, la persona integra lo que recuerda y decide cómo valorarlo. Si algunas solicitudes fueron resueltas rápidamente y otras presentaron retrasos, necesita un criterio para resumirlas. Un reactivo de frecuencia facilita una respuesta distinta a uno de satisfacción. La elección entre ambos depende de si interesa describir ocurrencia o evaluación. Formular una frecuencia con opciones de acuerdo introduce una tarea adicional que puede evitarse.

La selección final requiere trasladar ese juicio a las opciones disponibles. Las categorías deben ser reconocibles, pertinentes y suficientemente diferenciadas. Una persona sin experiencia en el proceso no encuentra una respuesta sustantiva adecuada en una escala que va de «muy insatisfecho» a «muy satisfecho». Si el sistema obliga a elegir, una respuesta completa puede ocultar la ausencia de información. Incorporar una ruta de no aplicabilidad conserva esa diferencia.

Como ejercicio didáctico, puede analizarse el reactivo «la institución siempre responde oportunamente». El término institución no identifica al actor, «siempre» introduce una exigencia absoluta y «oportunamente» admite criterios diferentes. Una alternativa para pilotar sería «en las últimas cuatro semanas, las solicitudes que realicé recibieron respuesta dentro del plazo informado». Esta versión todavía requiere

definir qué ocurre cuando no se comunicó un plazo y cuando la persona no presentó solicitudes.

Las entrevistas cognitivas permiten examinar estos pasos mediante preguntas sobre interpretación, recuerdo y elección. No se trata de comprobar si el participante conoce una respuesta correcta, sino de reconocer si la tarea que realiza coincide con la tarea prevista por el investigador. La revisión debe escuchar las explicaciones del informante antes de defender la redacción original. En ocasiones, una pregunta gramaticalmente sencilla exige información que el participante no posee.

1.2.2. Esfuerzo de respuesta y estrategias de simplificación

La teoría del satisficing explica que algunas personas pueden adoptar respuestas suficientes para continuar la encuesta sin completar el esfuerzo que demandaría una respuesta óptima. Krosnick relaciona estas estrategias con la dificultad de la tarea, la capacidad y la motivación del participante. El concepto no equivale a fraude: puede describir atajos ante una tarea exigente, incluso cuando la persona participa con una intención inicial de cooperación (Krosnick, 1991).

En un cuestionario administrativo, la simplificación puede aparecer como lectura parcial, selección de una categoría familiar o recuperación de un único episodio para responder varios reactivos. Estas conductas no se observan directamente a partir de un valor numérico. Los patrones de respuesta funcionan como señales que requieren interpretación. Una secuencia de puntuaciones iguales puede ser compatible con

una experiencia homogénea, especialmente si las preguntas están redactadas en la misma dirección.

La carga acumulada puede aumentar cuando se alternan temas sin transición o cuando cada pantalla cambia las reglas de respuesta. Un instrumento de veinte preguntas con escalas diferentes, instrucciones extensas y consultas a documentos puede resultar más exigente que otro con treinta preguntas concretas. Por ello, el análisis de simplificación debe atender a la complejidad de la tarea y a su organización, además de la longitud.

Gibson y Bowling estudiaron los efectos de la extensión mediante dos experimentos aleatorizados. Sus resultados aportan evidencia sobre la relación entre longitud y respuestas descuidadas, aunque no ofrecen una regla numérica universal para cualquier cuestionario. Esta evidencia respalda la evaluación de la extensión como una decisión de diseño y evita asumir que la atención se mantiene constante durante todo el recorrido (Gibson y Bowling, 2020).

El diseñador puede reducir esfuerzo innecesario al utilizar referentes concretos, mantener escalas coherentes y retirar preguntas repetidas que no añaden contenido. Sin embargo, no conviene eliminar toda repetición conceptual: una escala puede necesitar varios indicadores para representar un atributo. La cuestión consiste en distinguir diversidad de contenido y redundancia verbal. Reactivos casi idénticos pueden elevar la consistencia interna sin ampliar la información sobre el constructo.

La prevención tiene también una dimensión relacional. Explicar para qué servirá la encuesta y cuánto esfuerzo se espera permite una participación informada. Un mensaje que promete brevedad y conduce a numerosas pantallas puede

deteriorar la disposición a continuar. La expectativa comunicada debe basarse en el pilotaje y contemplar diferencias entre participantes. El tiempo anunciado es una estimación orientativa, no un plazo que la persona deba cumplir para que su opinión sea considerada válida.

1.2.3. Motivación, relevancia y carga percibida

La motivación para responder depende de la relación del participante con el tema y con la institución que solicita información. Un trabajador puede considerar útil una encuesta sobre dificultades de coordinación si espera que sus resultados se traduzcan en mejoras observables. También puede desconfiar de la confidencialidad o recordar experiencias anteriores en las que las respuestas no produjeron ninguna devolución. Estas condiciones modifican la experiencia de participación sin cambiar el número de preguntas.

La carga percibida expresa cómo se vive la tarea. Debe diferenciarse de la extensión objetiva y del tiempo registrado. Dos personas pueden invertir ocho minutos y experimentar esfuerzos distintos por su familiaridad con el tema, sus condiciones de lectura o las interrupciones que enfrentan. Una medición de carga aporta información sobre esa experiencia, pero necesita definir si pregunta por dificultad, cansancio, aburrimiento o esfuerzo.

Yu et al. examinan la carga más allá de la duración del cuestionario. Su investigación de 2026 destaca la relevancia de aspectos como repetición, desorganización y falta de sentido percibido. Este aporte refuerza la necesidad de preguntar cómo fue la experiencia de respuesta y de evitar que una única

medida temporal represente todo el concepto de carga (Yu et al., 2026).

La relevancia no debe convertirse en presión institucional. Una invitación enviada por una jefatura puede aumentar la participación y, al mismo tiempo, generar dudas sobre el uso de opiniones críticas. El diseño de la convocatoria debe facilitar una decisión voluntaria. La persona necesita saber quién tendrá acceso a los datos, cómo se comunicarán los resultados y qué información se utilizará para identificarla.

Como propuesta para el pilotaje, resulta útil separar una pregunta sobre dificultad de comprensión de otra sobre cansancio al finalizar. Si ambas se reúnen en «el cuestionario fue complicado, tedioso y cansado», una respuesta alta no permite identificar qué aspecto debe corregirse. La distinción facilita decisiones específicas: revisar vocabulario cuando existe dificultad, reorganizar bloques cuando aparece monotonía o retirar contenido accesorio cuando el esfuerzo acumulado resulta elevado.

La asociación entre carga y calidad puede depender de cómo se midieron ambas variables. Una pregunta redactada en términos de facilidad y luego interpretada como dificultad invierte el sentido de los valores. Del mismo modo, una medida de calidad que penaliza determinados patrones puede compartir información con una pregunta de atención. Antes de interpretar una relación, deben comprobarse las reglas de codificación y la independencia conceptual de los indicadores.

1.2.4. Dispositivos, presentación visual e interacción

La pantalla condiciona cómo se comprende y responde el cuestionario. Una cuadrícula que permite comparar varios

ítems en un computador puede exigir desplazamiento horizontal en un teléfono. Si las etiquetas de respuesta desaparecen durante el recorrido, el participante debe recordarlas. La presentación visual agrega así una demanda que no figura en el conteo de preguntas, pero que influye en el esfuerzo necesario para completar la tarea.

Vehovar et al. comparan diseños alternativos de preguntas en cuadrícula en computadores y dispositivos móviles mediante indicadores de calidad y estimaciones de encuesta. Su trabajo sitúa la decisión de formato dentro de una evaluación empírica. Cambiar una cuadrícula por ítems individuales no garantiza por sí solo una mejora uniforme: deben examinarse el contenido, el dispositivo y los resultados que se desean preservar (Vehovar et al., 2023).

La evaluación técnica puede comenzar por recorridos concretos: abrir el enlace, comprender las instrucciones, seleccionar una respuesta, corregirla y avanzar. Después debe comprobarse el comportamiento con distintos tamaños de pantalla y formas de interacción. La prueba realizada únicamente por quien programó el formulario suele pasar por alto dificultades que aparecen para una persona que no conoce la estructura prevista.

Las recomendaciones de accesibilidad del World Wide Web Consortium (W3C) ofrecen criterios sobre etiquetas, navegación por teclado, foco visible y presentación adaptable. Aplicarlas a una encuesta contribuye a reducir barreras de interacción. Una revisión visual aislada no certifica conformidad completa con la norma; la evaluación debe considerar los criterios aplicables y pruebas con personas que utilicen diferentes tecnologías de apoyo (W3C, 2024).

Décieux y Sischka analizan el comportamiento de respuesta entre teléfonos, tabletas y computadores en encuestas con diseño adaptable. Su comparación muestra la necesidad de estudiar los dispositivos dentro de condiciones de diseño concretas. Por ello, el dispositivo se incorpora al plan de pilotaje como una característica del contexto y no como una etiqueta automática de mayor o menor calidad (Décieux y Sischka, 2024).

Una solución administrativa proporcionada consiste en utilizar bloques temáticos breves y etiquetas completas en cada escala. Los mensajes de error deben explicar qué necesita corregirse y conservar lo que la persona ya respondió. El progreso mostrado debe corresponder al recorrido real, especialmente cuando existen saltos. Estos ajustes se verifican con tareas observables y se documentan como parte de la versión del instrumento.

1.2.5. Participación, confidencialidad y confianza en los resultados

La participación requiere condiciones que permitan responder con libertad. En una organización, las relaciones jerárquicas pueden afectar la forma en que se interpretan las preguntas y la disposición a expresar desacuerdo. Por ello, la convocatoria, el acceso al formulario y la presentación de resultados deben diseñarse considerando el vínculo entre participantes, investigadores y responsables de gestión. Una promesa de confidencialidad necesita corresponder a la configuración efectiva de la recolección.

Anonimato y confidencialidad tienen significados diferentes. Una encuesta anónima no permite vincular

razonablemente las respuestas con una persona. Una encuesta confidencial puede conservar esa vinculación, pero limita su acceso y uso. Solicitar correo institucional, cargo muy específico y departamento pequeño puede permitir identificar a alguien, aunque no se pregunte su nombre. La minimización de datos comienza al decidir qué variables de contexto son necesarias.

Tourangeau y Yan revisan las dificultades de preguntar por asuntos sensibles. Su análisis ayuda a comprender por qué el contenido y el contexto social pueden afectar la disposición a informar. En la investigación administrativa, una evaluación negativa de la jefatura o una descripción de conflictos puede resultar sensible aun cuando no involucre categorías habitualmente consideradas íntimas. La sensibilidad debe evaluarse desde la posición del participante (Tourangeau y Yan, 2007).

El seguimiento de la encuesta debe distinguir participación, finalización y respuesta a ítems. La American Association for Public Opinion Research (AAPOR) proporciona definiciones para documentar estados de los casos y tasas de resultado. El principio operativo es hacer explícito el denominador: quienes recibieron una invitación, quienes eran elegibles o quienes iniciaron no representan necesariamente el mismo conjunto. Cuando se difunde un enlace abierto, puede desconocerse el total de personas que tuvieron oportunidad real de participar (AAPOR, 2023).

La devolución institucional cierra el proceso de participación. Un informe útil explica qué se encontró, con qué alcance y qué acciones se consideran. También reconoce aquello que la encuesta no permite decidir. Publicar rankings de unidades pequeñas sin atender al número y perfil de

participantes puede producir interpretaciones desproporcionadas. Las decisiones sobre presentación deben definirse junto con la finalidad del estudio y las condiciones de protección acordadas.

La confianza se fortalece cuando el equipo puede mostrar la secuencia que une conceptos, preguntas, datos y conclusiones. Esto exige conservar versiones, justificar transformaciones y explicar la participación observada. Una encuesta eficiente permite responder la pregunta que la originó y deja evidencia suficiente para revisar cómo se alcanzó esa respuesta. Los siguientes capítulos desarrollan las variables y los procedimientos necesarios para construir esa continuidad.

CAPÍTULO II

EXTENSIÓN, TIEMPO DE RESPUESTA Y CALIDAD DE LOS DATOS

La relación entre extensión y calidad solo puede estudiarse con claridad cuando sus variables están bien definidas. El número de preguntas, la duración registrada y la carga percibida describen aspectos relacionados, pero diferentes. Del mismo modo, la calidad de respuesta reúne señales que pueden corresponder a problemas distintos. Tratar todas estas medidas como equivalentes dificulta interpretar los resultados y conduce a recomendaciones poco específicas.

Este capítulo presenta definiciones operativas para las variables centrales del libro. Explica cómo observarlas en encuestas administrativas y qué información adicional requiere cada indicador. El propósito es construir una base conceptual que permita comprender el estudio y diseñar instrumentos futuros con reglas explícitas de medición.

2.1. Dimensiones de la extensión y del esfuerzo solicitado

El diseño establece una cantidad potencial de contenido, mientras que la experiencia de cada participante depende del recorrido que efectivamente realiza. La medición debe distinguir ambas perspectivas para no atribuir a la longitud nominal todo lo que sucede durante la respuesta.

2.1.1. Extensión prevista, exposición efectiva y respuestas obtenidas

La extensión prevista es el número de reactivos incluidos en una versión del cuestionario. Debe contarse con una regla estable. Una cuadrícula con diez afirmaciones contiene diez tareas de respuesta, aunque la plataforma la denomine una

pregunta. Un bloque de selección múltiple puede exigir varias decisiones dentro de un mismo campo. Por ello, el inventario necesita describir el tipo de tarea además del número de elementos.

La exposición efectiva corresponde a los reactivos que el sistema presentó a una persona. Los saltos, filtros y módulos opcionales pueden hacer que dos participantes de la misma versión reciban cantidades diferentes. Si una persona no utilizó determinado servicio, omitir el bloque correspondiente evita solicitar juicios sin experiencia. Esa ausencia es parte del diseño y no debe codificarse como omisión accidental.

La cantidad respondida indica cuántos reactivos recibieron una respuesta. Su diferencia con la exposición permite estudiar omisiones, siempre que se separe la decisión de no contestar de los campos no mostrados. Una base que conserva únicamente los valores finales y no registra los recorridos dificulta esa distinción. El diccionario de datos debe definir cómo se representan preguntas omitidas por salto, respuestas no sustantivas y celdas vacías.

En un ejemplo ficticio, un cuestionario tiene cuarenta reactivos previstos. Una trabajadora recibe treinta porque no utiliza un módulo de compras y responde veintinueve. Su exposición es treinta y tiene una omisión entre los reactivos mostrados. Calcular una tasa con cuarenta como denominador daría una impresión errónea de su comportamiento. Tampoco sería correcto atribuirle once omisiones, pues diez preguntas no fueron pertinentes para su recorrido.

La longitud nominal puede incluir consentimiento, elegibilidad y variables de contexto, o reservarse para los reactivos sustantivos. Ninguna convención debe quedar implícita. El estudio de base informa versiones de treinta,

cuarenta, cincuenta y sesenta y cinco reactivos; el libro conserva esas denominaciones. El capítulo tercero detalla la construcción de las cuatro versiones utilizadas en la investigación.

Para comparar versiones, conviene mantener un núcleo común y registrar la posición de sus reactivos. Así puede distinguirse si cambia la respuesta a una misma pregunta cuando se añade contenido. Comparar indicadores calculados sobre conjuntos completamente diferentes mezclas efectos de longitud con diferencias de contenido. La extensión se vuelve interpretable cuando está acompañada por un mapa de las tareas y de su ubicación.

2.1.2. Complejidad de los reactivos y cobertura conceptual

La complejidad es la demanda que una pregunta impone para ser comprendida y respondida. Puede provenir de la redacción, de la recuperación de información, del juicio solicitado o de la interacción digital. Una pregunta breve puede exigir consultar registros de varios meses; otra más extensa puede describir una situación muy concreta. El conteo de palabras aporta una señal de forma, pero no reemplaza la evaluación de la tarea.

La cobertura conceptual se refiere a la representación de las dimensiones relevantes del constructo. En una medida de coordinación, pueden importar la claridad de responsables, el intercambio de información y el seguimiento de acuerdos. Un conjunto de preguntas centrado solo en reuniones puede dejar fuera otros componentes. Reducir longitud sin revisar esa cobertura puede mejorar la experiencia de respuesta y empobrecer la interpretación del resultado.

Clark y Watson destacan que la construcción de instrumentos requiere delimitar el dominio conceptual y desarrollar un conjunto de ítems que permita examinarlo. La selección posterior debe atender a la estructura y al contenido de la medición. Este planteamiento ayuda a evitar que la depuración se limite a retirar reactivos que disminuyen un coeficiente estadístico, dejando una escala estrecha o redundante (Clark y Watson, 2019).

Un procedimiento didáctico consiste en asignar a cada reactivo una descripción de la experiencia que solicita. Si varias preguntas remiten a la misma situación y usan términos apenas distintos, existe una posible redundancia. Si una dimensión aparece en una única pregunta muy general, puede estar insuficientemente representada. La matriz de contenido hace visibles ambas situaciones antes de que la longitud quede fijada.

La complejidad también aumenta con preguntas dobles. «El procedimiento es claro y rápido» reúne comprensión y duración. Una persona puede evaluar favorablemente la claridad y negativamente la rapidez, sin disponer de una respuesta que represente ambas experiencias. Dividir la pregunta eleva el conteo, pero mejora la diferenciación conceptual. La decisión de añadir un reactivo puede ser eficiente si elimina una ambigüedad importante para el análisis.

El diseño debe establecer prioridades de información. Los contenidos esenciales responden directamente a la pregunta principal; los complementarios ayudan a interpretar diferencias; los exploratorios abren asuntos nuevos. Esta clasificación permite negociar reducciones sin tratar todas las preguntas como igualmente necesarias. Los módulos

exploratorios pueden administrarse en otra fase cuando su inclusión amenaza el funcionamiento del núcleo principal.

2.1.3. Tiempo total, tiempo de interacción y pausas

El tiempo total de una sesión suele calcularse como la diferencia entre una marca de inicio y otra de finalización. Antes de utilizarla, debe comprobarse qué evento produce cada marca: abrir el formulario, entrar a una página, comenzar una respuesta o enviar el cuestionario. La definición depende del sistema de recolección. Dos plataformas pueden generar variables con el mismo nombre que representan momentos distintos.

El tiempo de interacción busca aproximar el período dedicado a responder, excluyendo algunas pausas cuando existen registros apropiados. Esa distinción requiere reglas técnicas y conceptuales. La pérdida de foco de una ventana puede indicar que la persona abrió otra aplicación, pero también que consultó información necesaria. Ningún registro de navegación revela por sí solo el contenido de la actividad ni la atención mental del participante.

Décieux estudia la multitarea secuencial dentro del dispositivo mediante señales de cambio de foco. Su investigación encuentra asociaciones con indicadores de calidad de respuesta. El aporte pertinente para este libro es metodológico: una duración prolongada puede contener interrupciones y necesita información adicional para interpretarse. No corresponde clasificar todo tiempo largo como distracción ni toda pausa como incumplimiento (Décieux, 2024).

La duración por ítem puede calcularse dividiendo el tiempo observado por la cantidad de tareas respondidas, pero su significado depende de la homogeneidad del contenido. Una respuesta abierta y una selección de frecuencia no demandan el mismo trabajo. En cuestionarios con módulos distintos, resulta más útil resumir tiempos por bloque y dispositivo que establecer un único límite de velocidad para todo el instrumento.

Un ejemplo ficticio permite apreciar el problema. Dos personas tardan doce minutos. La primera responde sin interrupciones; la segunda realiza una pausa de cinco minutos para atender una tarea laboral. Si el sistema solo registra inicio y envío, ambas aparecen con la misma duración. El dato es correcto como tiempo transcurrido, pero no permite afirmar que ambas dedicaron doce minutos de atención continua. La interpretación debe conservar esa limitación.

Para describir tiempos asimétricos conviene informar mediana y percentiles, además de la media cuando resulte útil. Los valores negativos o cronologías imposibles requieren revisión técnica. Los tiempos extremos plausibles deben conservarse inicialmente y examinarse con criterios definidos. Su eliminación automática puede retirar a personas que utilizaron apoyos de accesibilidad, consultaron documentación o respondieron con especial detenimiento.

2.1.4. Carga objetiva y experiencia subjetiva de respuesta

La carga objetiva describe demandas identificables del instrumento, como cantidad de tareas, longitud del texto, número de pantallas, consultas a documentos o cambios de escala. La carga percibida expresa la valoración que el

participante hace de esas demandas. Ambas pueden relacionarse, pero no son equivalentes. Una encuesta larga sobre un asunto familiar puede ser vivida de manera diferente a una encuesta breve sobre información difícil de recordar.

Kunz y Gummer investigan la relación entre carga objetiva, carga percibida y calidad de respuesta. Su trabajo muestra la utilidad de medir la percepción durante el recorrido y de no asumir que la posición de una pregunta resume por completo la carga experimentada. En el diseño administrativo, esta distinción permite formular intervenciones específicas sobre contenido, secuencia e interacción (Kunz y Gummer, 2025).

La medición debe elegir un referente claro. «¿Qué dificultad tuvo para comprender las preguntas?» se centra en comprensión. «¿Cuánto cansancio le produjo completar el cuestionario?» pregunta por una experiencia de fatiga. «¿Cuánto esfuerzo dedicó a responder?» se aproxima a la inversión de recursos. Estas preguntas pueden relacionarse, pero no deberían interpretarse como sinónimos sin evidencia que sostenga una dimensión compartida.

El sentido de la escala necesita documentación. Cuando cinco significa «muy difícil», los valores altos representan mayor dificultad. Si cinco significa «muy sencillo», los valores altos representan facilidad. Transformar una escala en otra exige una recodificación explícita. La descripción del indicador, el texto que vio el participante y el código analítico deben coincidir; de lo contrario, puede invertirse la interpretación de una asociación.

En el estudio de base aparecen formulaciones diferentes para la carga, incluyendo facilidad y dificultad. Por ello, la lectura de sus coeficientes debe reconocer que la codificación exacta necesita corroboración con el formulario aplicado y la

sintaxis original. Los capítulos tercero y cuarto desarrollan esta cuestión. La precaución no impide utilizar el estudio, pero delimita qué recomendaciones pueden sostenerse a partir de sus tablas.

En el pilotaje, una breve conversación posterior puede ayudar a interpretar valores altos. Si las personas describen una pregunta como difícil porque mezcla varias situaciones, se requiere revisar el reactivo. Si mencionan falta de acceso al dato, puede ser necesario cambiar el informante o la fuente. La medición de carga resulta útil cuando orienta una decisión de diseño, y no únicamente cuando se incorpora como otra variable a una regresión.

2.1.5. Relaciones entre longitud, duración y calidad

Una hipótesis razonable es que añadir contenido incremente el tiempo requerido y que, bajo ciertas condiciones, ese incremento se acompañe de cambios en la atención. Sin embargo, el contenido añadido también puede alterar la relevancia, la dificultad y el interés. Estas rutas explican por qué la relación entre longitud y calidad no debe reducirse a una cadena automática. El diseño de la comparación determina qué interpretaciones son posibles.

Galesic y Bosnjak estudiaron la longitud del cuestionario en relación con participación e indicadores de calidad en una encuesta web. Este antecedente contribuye a distinguir expectativas sobre duración y extensión efectivamente administrada. La decisión de comenzar y la conducta durante el cuestionario ocurren en momentos diferentes; por tanto, el efecto de la longitud puede expresarse tanto en la participación como en el proceso de respuesta (Galesic y Bosnjak, 2009).

Cernat et al. compararon diseños de encuesta, entre ellos un cuestionario más corto organizado mediante un diseño matricial. Encontraron calidad semejante en varios aspectos, con diferencias en otros indicadores y niveles de equivalencia de medición. Sus resultados impiden presentar la literatura como una demostración uniforme de que todo cuestionario más largo produce datos peores. Importan el diseño completo y el criterio con el que se define calidad (Cernat et al., 2024).

La relación entre tiempo y calidad puede ser no lineal. Respuestas extremadamente rápidas podrían indicar lectura insuficiente, mientras que tiempos largos pueden reflejar consultas, dificultad o pausas. Un modelo cuadrático ofrece una forma de explorar curvatura, pero también puede ajustarse de manera inadecuada si existen valores extremos o restricciones injustificadas. La forma matemática necesita contrastarse con la distribución y con mecanismos plausibles.

La carga percibida puede actuar como resultado de la longitud y, a la vez, relacionarse con la respuesta. Incluirla como covariable en un modelo del efecto total de la extensión cambia la pregunta analítica si se encuentra en una ruta intermedia. El plan debe distinguir entre estimar una diferencia global entre versiones y examinar posibles mecanismos. Llamar «variable de control» a una medida no explica qué función cumple en el modelo.

En una propuesta aplicada, conviene formular hipótesis diferenciadas: comparación de versiones en indicadores comunes, relación de la carga con problemas identificados y descripción de tiempos por recorrido. La edad o el dispositivo pueden utilizarse para examinar heterogeneidad cuando exista justificación y tamaño suficiente. No deben convertirse en criterios para asignar calidad a las personas de antemano. El

interés está en comprender condiciones de medición que puedan mejorarse.

2.2. Indicadores y límites de la calidad de respuesta

Los indicadores de calidad aproximan comportamientos o propiedades de los datos. Su utilidad depende de que la señal observada tenga una relación justificada con el problema que se desea detectar. La revisión debe distinguir problemas de participación, dificultades de medición y patrones que solamente requieren una comprobación adicional.

2.2.1. Abandono, omisión y respuestas no sustantivas

El abandono se refiere a una interrupción antes del punto definido como finalización. Para medirlo se necesita conocer cuántas personas iniciaron y qué evento se considera un inicio. Si el sistema guarda únicamente cuestionarios enviados, la ausencia de registros incompletos no demuestra ausencia de abandono. Debe verificarse si existen datos sobre sesiones iniciadas, accesos o formularios parciales y si su tratamiento corresponde a la información proporcionada al participante.

Una tasa de abandono entre iniciadores puede expresarse como el número de sesiones elegibles que no finalizaron dividido por el número de sesiones elegibles iniciadas, multiplicado por cien. Esta tasa no equivale a una tasa de respuesta sobre invitaciones. En un ejemplo ficticio, de doscientas invitaciones se producen cien inicios y ochenta finalizaciones. El abandono entre iniciadores es veinte por ciento; la finalización sobre invitaciones es cuarenta por ciento. Cada cifra responde a una pregunta diferente.

La omisión de un ítem ocurre cuando una persona expuesta a la pregunta no proporciona respuesta. Puede medirse por pregunta, por participante o por bloque. El porcentaje de participantes con al menos una omisión aumenta su oportunidad de ocurrencia cuando se presentan más reactivos. Por ello, comparar ese indicador entre versiones de distinta longitud requiere cautela y debería acompañarse con tasas ajustadas por exposición.

Las respuestas «no sabe», «no aplica» y «prefiere no responder» contienen información distinta. La primera puede indicar falta de conocimiento; la segunda, ausencia de experiencia relevante; la tercera, una decisión de reserva. Agruparlas en un único código dificulta identificar qué debe corregirse. Su interpretación también depende de si se ofrecieron de manera visible o si fueron generadas posteriormente por el sistema.

Sischka et al. encontraron que la respuesta forzada podía aumentar la reactancia y el abandono, además de afectar la extensión de respuestas abiertas. Esta evidencia cuestiona la idea de que hacer obligatorios todos los campos asegura mejores datos. Esta evidencia respalda que las respuestas sustantivas puedan omitirse y que las decisiones de consentimiento y elegibilidad se gestionen como condiciones del recorrido, con una salida clara (Sischka et al., 2022).

El tratamiento posterior debe conservar la procedencia del dato ausente. Una imputación puede ser apropiada bajo determinados supuestos analíticos, pero no elimina la necesidad de conocer por qué falta información. En estudios sobre calidad de respuesta, imputar antes de describir las omisiones puede borrar precisamente el fenómeno que se

intenta estudiar. La base original y las versiones destinadas al análisis deben mantenerse diferenciadas y documentadas.

2.2.2. Respuestas lineales, extremas y de punto medio

La respuesta lineal o straight-lining describe una secuencia de categorías iguales, habitualmente dentro de un bloque con escala común. Puede detectarse mediante variación nula o mediante la longitud de una secuencia repetida. Su significado depende del contenido. En preguntas que describen experiencias realmente semejantes, una serie de respuestas iguales puede ser sustantivamente coherente; en ítems diversos, puede justificar una revisión adicional.

El indicador debe especificar qué reactivos se comparan y cuántos se requieren. Calcular variación sobre un conjunto que mezcla edad, escalas de satisfacción y códigos de sector carece de sentido. También es necesario considerar que algunos participantes pueden responder pocos ítems por los filtros. Un patrón de tres respuestas iguales no aporta la misma evidencia que una secuencia extensa que atraviesa contenidos heterogéneos.

Las respuestas extremas utilizan los límites de la escala. Pueden reflejar experiencias intensas, estilos de respuesta o lectura insuficiente. La selección del punto medio puede expresar neutralidad, una experiencia intermedia o dificultad para elegir. Ninguna de estas posibilidades se distingue automáticamente contando categorías. El diseño debe permitir separar la neutralidad sustantiva de la falta de conocimiento cuando ambas sean pertinentes.

Goldammer et al. compararon indicadores para detectar respuestas descuidadas y examinaron sus consecuencias en

medidas organizacionales. Encontraron diferencias de desempeño entre procedimientos, lo que respalda la necesidad de utilizar evidencia combinada y contextual. El patrón repetido resulta más informativo cuando coincide con otras señales y con un contenido que hace difícil sostener una interpretación sustantiva razonable (Goldammer et al., 2020).

El índice del estudio de base incorpora una proporción de respuestas en las categorías uno, tres y cinco. Esa regla reúne extremos y punto medio. Al penalizar tres de cinco opciones, requiere una justificación específica para distinguir estilo de respuesta y contenido legítimo. El libro conserva la regla al describir el estudio, sin presentarla como una forma universalmente validada de medir error.

En la práctica, conviene producir un indicador de revisión por bloque y conservar las respuestas originales. La decisión sobre un registro debe documentar qué señales confluyeron y qué alternativas se consideraron. Si se elimina toda respuesta uniforme, puede reducirse artificialmente la presencia de experiencias muy favorables o desfavorables. La calidad del análisis incluye evitar que los procedimientos de depuración impongan una distribución deseada a las opiniones.

2.2.3. Atención, consistencia de pares y participación fraudulenta

La atención puede evaluarse mediante autoinforme, instrucciones específicas, preguntas de conocimiento pertinente o señales del proceso. Cada método tiene limitaciones. Una persona puede declarar alta atención por deseabilidad social; una instrucción inesperada puede ser pasada por alto a pesar de una participación generalmente

cuidadosa. La elección depende del propósito y de la carga adicional que el control introduce en el cuestionario.

Los pares de reactivos requieren una evaluación semántica especialmente cuidadosa. Dos preguntas con palabras opuestas no siempre representan negaciones lógicas. «Ayudar a cambio de esfuerzo» y «evitar ayudar a quien se esfuerza» contienen condiciones que pueden interpretarse de manera distinta. Responder con la misma categoría no demuestra necesariamente contradicción. Las entrevistas cognitivas deben examinar cómo se comprenden esos pares antes de emplearlos como criterio de atención.

Existe además un problema de exposición. Si una versión tiene más oportunidades de fallar un control, la probabilidad de presentar al menos una señal puede aumentar, aunque el comportamiento por control sea idéntico. Como ilustración matemática ficticia, si cada control independiente tiene probabilidad de señal de 0,05, la probabilidad de al menos una señal es aproximadamente 0,185 con cuatro controles y 0,537 con quince. La independencia y la probabilidad constante son supuestos del ejemplo, no propiedades comprobadas del estudio.

Meade y Craig presentan procedimientos para identificar respuestas descuidadas, mientras que Curran revisa métodos y decisiones de detección. En conjunto, estos trabajos permiten tratar la evaluación como un problema de medición que necesita múltiples fuentes de información. Un criterio de exclusión no queda validado por el hecho de producir una base más consistente después de retirar casos (Curran, 2016; Meade y Craig, 2012).

La participación fraudulenta implica otro problema: registros que no corresponden a la participación prevista,

identidades falsas u otras formas de manipulación. Lawlor et al. proponen el marco REAL para abordar este fenómeno durante la recolección. Nur et al. describen experiencias de detección y manejo del fraude en un proyecto específico. Sus hallazgos justifican preparar un procedimiento, pero sus tasas no deben trasladarse como prevalencias esperables para toda encuesta administrativa (Lawlor et al., 2021; Nur et al., 2024).

Compartir una dirección de red no prueba que dos respuestas pertenezcan a la misma persona. En instituciones, varias personas pueden utilizar una conexión común. De manera similar, una respuesta rápida puede provenir de familiaridad con el contenido. La revisión debe evitar conclusiones automáticas y conservar un registro de razones. La prevención de duplicaciones se diseña junto con las condiciones de acceso y de privacidad, evitando recolectar información identificable que no resulte necesaria.

2.2.4. Fiabilidad, estructura de medición y valores atípicos

La fiabilidad describe la consistencia o precisión de una puntuación bajo un modelo y condiciones determinados. El alfa de Cronbach es un indicador de consistencia interna, pero su interpretación depende de supuestos sobre los reactivos y la estructura de la medición. Un valor alto puede acompañar preguntas redundantes. Tampoco permite establecer que todas las personas respondieron con atención ni que la escala mide el constructo que su nombre anuncia.

McNeish examina las limitaciones del uso rutinario de alfa y alternativas de estimación. El aporte práctico consiste en elegir el coeficiente de acuerdo con el modelo, en lugar de tratar

un umbral único como certificado general de calidad. Cuando se dispone de un modelo factorial adecuado, otros estimadores, como omega, pueden aportar información pertinente, siempre que también se expliciten sus supuestos (McNeish, 2018).

La estructura interna se estudia observando cómo se relacionan los reactivos y si esas relaciones corresponden a las dimensiones previstas. Una escala de coordinación y otra de apoyo pueden estar relacionadas sin representar una sola dimensión. Sumar ambas exige justificar qué significa la puntuación conjunta. La evidencia estadística debe contrastarse con la definición conceptual; un modelo ajustado sobre una muestra no autoriza cualquier uso de sus resultados.

Los valores atípicos describen observaciones poco frecuentes respecto de una distribución o un patrón multivariado. La distancia de Mahalanobis considera la combinación de variables y sus relaciones, pero necesita una matriz de covarianza estimable y una decisión sobre el conjunto de indicadores. Una observación distante puede ser una experiencia válida y relevante. En una organización heterogénea, retirar a todas las personas con condiciones poco habituales puede ocultar problemas sustantivos.

Wind y Lugu exploran procedimientos de teoría de respuesta al ítem para estudiar calidad de datos. Su trabajo amplía las opciones de evaluación, pero no implica que todo instrumento administrativo requiera modelos complejos. La selección del procedimiento debe corresponder a la escala, el tamaño de la muestra, los objetivos y la capacidad de interpretar las señales producidas (Wind y Lugu, 2024).

La fiabilidad debe estimarse sobre conjuntos conceptualmente coherentes y reportarse con suficiente

información del análisis. Conviene indicar número de reactivos, tamaño de muestra utilizado, tratamiento de omisiones y justificación del coeficiente. Los valores atípicos se registran como señales y se examinan mediante análisis de sensibilidad. Ninguna de estas operaciones sustituye la evaluación de contenido, del proceso de respuesta y de las condiciones en que se obtuvieron los datos.

2.2.5. Índices compuestos y evaluación mediante varios indicadores

Un índice compuesto integra diferentes componentes en una puntuación. Puede facilitar comparaciones, pero introduce decisiones sobre selección, dirección, escala y ponderación. Si los componentes representan problemas distintos, el índice funciona como una construcción analítica que resume esas decisiones. No debe asumirse que constituye una variable latente unidimensional ni que sus ponderaciones son naturales por el hecho de haber asignado el mismo peso a cada elemento.

El estudio de base define calidad de respuesta en una escala de cero a siete restando siete penalizaciones: abandono, presencia de omisiones, variación nula, proporción de categorías seleccionadas, falta de atención declarada, falta de atención observada y valores atípicos. Para describirlo de manera compacta, se representa cada penalización con una letra entre a y g. Los valores más altos del resultado indican menor penalización bajo esa regla particular.

La suma requiere que los componentes estén expresados entre cero y uno. Algunos son binarios y otros son proporciones o promedios reescalados. Asignarles igual rango no garantiza que representen problemas de igual importancia.

Además, una misma dificultad puede afectar varios componentes y recibir penalizaciones repetidas. El tratamiento de una sesión sin respuestas resulta especialmente importante porque algunas señales son imposibles de calcular y otras sí son observables.

Tabla 2.
Señales de calidad y límites de interpretación

Señal	Información que aporta	Límite de interpretación
Abandono	Interrupción del recorrido	Necesita registro de inicios
Omisión	Falta de respuesta en ítems expuestos	Puede reflejar reserva o falta de experiencia
Patrón uniforme	Repetición de categorías	Puede ser sustantivamente coherente
Control de atención	Respuesta ante una tarea específica	Depende de comprensión y exposición
Tiempo inusual	Duración diferente a la referencia	No identifica por sí solo la causa
Valor atípico	Perfil estadísticamente infrecuente	No equivale a error

Fuente: elaboración propia.

Bhaktha et al. proponen en 2026 un marco que diferencia estilos de respuesta, esfuerzo insuficiente y manipulación, vinculándolos con indicadores observables. Su contribución refuerza una distinción central: la señal estadística no identifica de manera inequívoca el proceso que la produjo. Por ello, los indicadores del estudio se examinan de forma conjunta y se interpretan dentro de su contexto (Bhaktha et al., 2026).

Un tablero permite observar componentes que evolucionan en direcciones distintas. Si una versión reduce omisiones y aumenta abandono, el equipo necesita conocer ambos resultados. Un índice adicional requiere justificar su propósito y comprobar la sensibilidad a las ponderaciones. Publicar los componentes y las reglas permite revisar la interpretación.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN SOBRE EXTENSIÓN DEL CUESTIONARIO Y CALIDAD DE RESPUESTA

La investigación examinó la relación entre la extensión de un cuestionario digital y la calidad de las respuestas obtenidas de trabajadores de los sectores público y privado de Ecuador. Para ello, se aplicaron cuatro versiones de un instrumento adaptado del Multifactor Leadership Questionnaire, con 30, 40, 50 y 65 reactivos. El análisis se centró en el comportamiento de respuesta y en sus indicadores estadísticos. El contenido sobre liderazgo proporcionó el contexto de las preguntas, mientras que el objeto de estudio fue la calidad de la información producida al contestarlas.

El trabajo se desarrolló mediante un enfoque cuantitativo. Se utilizaron cuestionarios estructurados, registros de duración y procedimientos estadísticos para comparar las versiones y estudiar las asociaciones dentro de cada grupo. La muestra analizada reunió 473 participantes y el trabajo de campo se extendió desde noviembre de 2024 hasta diciembre de 2025. Este capítulo presenta las decisiones metodológicas que dieron origen a los resultados, desde la definición del diseño hasta la construcción del índice de calidad de respuesta.

3.1. Enfoque de investigación y organización del estudio

3.1.1. Problema de investigación y objetivos del análisis

El problema que orientó el estudio fue determinar si aumentar la cantidad de preguntas de una encuesta digital se relacionaba con cambios en la calidad de las respuestas. En los estudios administrativos, la facilidad de incorporar reactivos a una plataforma puede llevar a formularios extensos, pero la disponibilidad de más respuestas no asegura una mejor medición. Era necesario observar qué sucedía con la atención,

las omisiones, la consistencia y otros indicadores cuando los participantes respondían versiones de diferente longitud.

El objetivo general consistió en determinar el impacto de la extensión del cuestionario sobre la calidad de respuesta en encuestas digitales para estudios de ciencias administrativas. Este propósito se desarrolló mediante cuatro objetivos específicos: identificar las dimensiones observables de la calidad de respuesta, comparar cuestionarios de distinta extensión, examinar las variables relacionadas con la calidad e investigar las condiciones de extensión y duración asociadas con mejores puntuaciones. La evaluación se concentró en las versiones efectivamente aplicadas y en la población participante.

La hipótesis nula de la comparación principal estableció que la calidad media de respuesta era igual entre los cuatro grupos. Su contraste permitió evaluar si las diferencias observadas podían atribuirse únicamente a la variabilidad muestral bajo el modelo estadístico utilizado. En el análisis de cada versión se examinó, además, la relación de la calidad con el tiempo de llenado, la carga percibida, la edad y el género. El tiempo se incorporó mediante un término lineal y otro cuadrático para investigar una posible relación no lineal.

La identificación de una duración óptima constituyó una pregunta de investigación. Su respuesta exigía distinguir entre el tiempo promedio realmente observado y el punto de máximo de una curva estimada. Esta distinción acompañó el análisis porque una duración media describe a los participantes, mientras que un máximo matemático depende de la especificación del modelo, de la precisión de sus coeficientes y del rango de tiempos disponible.

3.1.2. Enfoque cuantitativo y lógica hipotético deductiva

El estudio adoptó una fundamentación positivista y una lógica hipotético deductiva. A partir de antecedentes sobre extensión de cuestionarios y calidad de respuesta, se formularon hipótesis susceptibles de contrastación con observaciones cuantificables. Las respuestas se codificaron numéricamente y se transformaron en variables para el análisis descriptivo e inferencial. Esta relación entre objetivos, medición y contraste estadístico corresponde al enfoque cuantitativo expuesto por Creswell y Creswell (2018).

La investigación tuvo una finalidad aplicada, pues buscó aportar evidencia para mejorar el diseño de encuestas digitales utilizadas en ciencias administrativas. Su contribución esperada no consistió en caracterizar los estilos de liderazgo de los trabajadores ecuatorianos, sino en conocer cómo se comportaban los indicadores de calidad cuando cambiaba la longitud de un formulario. Esta delimitación explica que el análisis se concentrara en las condiciones de respuesta y no en construir perfiles organizacionales a partir de las puntuaciones de liderazgo.

La técnica de recolección fue la encuesta digital autoadministrada. La información sobre atención y carga se obtuvo mediante reactivos estructurados incluidos en el propio cuestionario; el tiempo se calculó a partir de sus registros de inicio y finalización. La valoración experta y el pilotaje formaron parte de la revisión del instrumento. No constituyeron entrevistas a los participantes ni una fase de análisis cualitativo.

El interés explicativo del estudio se expresó en la comparación de condiciones de extensión y en la estimación de relaciones estadísticas. Sin embargo, la fuerza de una

conclusión causal depende también de cómo se forman los grupos y de las condiciones de aplicación. Por ello, el análisis mantuvo la finalidad de estudiar el efecto de la extensión, pero interpretó los resultados considerando el reclutamiento voluntario y la aplicación de las versiones en períodos distintos.

3.1.3. Diseño de comparación de cuatro versiones

La variable manipulada fue la extensión del cuestionario, representada por cuatro cantidades de reactivos. El instrumento de 30 preguntas constituyó el grupo de referencia, denominado grupo de control en el estudio. Las versiones de 40, 50 y 65 preguntas correspondieron a los grupos experimentales I, II y III. Cada grupo respondió una versión y las comparaciones se organizaron como análisis entre grupos independientes.

Tabla 3.
Condiciones de extensión del cuestionario

Grupo	Preguntas	Función en el estudio
Control	30	Versión de referencia
Experimental I	40	Primera ampliación
Experimental II	50	Segunda ampliación
Experimental III	65	Versión más extensa

Fuente: Elaboración propia.

Las versiones compartieron una base de contenido procedente del MLQ, una escala de respuesta de cinco categorías y la inclusión de reactivos para observar atención y consistencia. El cuestionario de 65 preguntas fue la versión más

extensa y las formas de 50, 40 y 30 preguntas se obtuvieron mediante la supresión de reactivos. Este procedimiento permitió estudiar una variación deliberada de la longitud dentro de una misma familia de instrumentos.

El reclutamiento combinó conveniencia y bola de nieve. Cada versión se difundió en un período distinto y la participación fue voluntaria. Estas condiciones no documentan una asignación aleatoria de personas a las cuatro versiones. La ausencia de selección discrecional de cada participante tampoco equivale, por sí misma, a aleatorización. Por tanto, se trató de una comparación con manipulación de la extensión, cuyo alcance debía considerar posibles diferencias entre grupos y momentos de recolección.

La denominación de grupo de control identifica aquí la versión utilizada como referencia. No implica que ese grupo estuviera libre de omisiones, desatención o respuestas inconsistentes. Precisamente, todos esos comportamientos se midieron también en el cuestionario de 30 preguntas. De igual manera, modificar la cantidad de reactivos alteró el conjunto de preguntas y controles presentes, aspecto relevante al interpretar la comparabilidad de los indicadores.

3.1.4. Población y procedimiento de muestreo

La población de interés estuvo constituida por trabajadores de los sectores público y privado de Ecuador. La unidad de análisis fue la persona que respondió el cuestionario digital. La experiencia laboral proporcionó un marco pertinente para contestar afirmaciones relacionadas con liderazgo, mientras que la modalidad en línea permitió observar el llenado autoadministrado del formulario.

La difusión comenzó a través de los círculos de contacto del investigador. Posteriormente, se solicitó a quienes participaban que compartieran el enlace con otras personas, lo que amplió el reclutamiento mediante referencias sucesivas. El procedimiento correspondió a un muestreo no probabilístico por conveniencia y bola de nieve. No se dispuso de un marco muestral para seleccionar aleatoriamente a los trabajadores ni se calcularon probabilidades individuales de inclusión.

Esta estrategia facilitó la obtención de respuestas, aunque vinculó la composición de los grupos con las redes de contacto, la disponibilidad para participar y el acceso a la encuesta. Tales condiciones son importantes incluso cuando el interés recae en el proceso de respuesta: las personas que aceptan una invitación pueden diferir de quienes no la reciben o no contestan en aspectos relacionados con motivación y familiaridad digital.

En consecuencia, la muestra permitió estudiar asociaciones y diferencias en los grupos participantes, pero no estimar la prevalencia nacional de respuestas de baja calidad. La pertenencia a los sectores público y privado delimitó la población de interés; el análisis presentado no estableció una comparación de resultados entre esos dos sectores. Atribuir diferencias sectoriales habría requerido información y contrastes específicos que no formaron parte de los resultados expuestos.

3.1.5. Tamaño de muestra y período de recolección

La planificación del tamaño de muestra se basó en una fórmula de estimación de una media para una población considerada infinita. Se emplearon un valor Z de 1,96,

correspondiente a un nivel nominal de confianza del 95 %, una varianza de referencia de 0,3729 y una media de 4,96. Los dos últimos valores procedieron del antecedente de Izaguirre Olmedo et al. (2025). El error admitido se definió como el 3 % de esa media, equivalente a 0,1488 unidades.

$$n = \frac{1{,}96^2 \times 0{,}3729}{(0{,}03 \times 4{,}96)^2} = 64{,}80 \approx 65$$

El cálculo produjo un valor aproximado de 64,80 participantes, que se redondeó a 65. El investigador elevó la meta operativa a un mínimo de 80 personas por grupo. Esta decisión aumentó el número de observaciones previsto para las comparaciones, aunque el cálculo utilizado correspondió a precisión de una media y no a una estimación específica de potencia para detectar diferencias entre las cuatro versiones.

La muestra analizada superó esa meta en todos los grupos: 186 participantes contestaron la versión de 30 preguntas, 92 la de 40, 84 la de 50 y 111 la de 65. La suma fue de 473 participantes. El tamaño desigual de los grupos se conservó en el análisis, en lugar de reducir artificialmente el grupo más numeroso para igualarlo con los restantes.

El trabajo de campo se realizó entre noviembre de 2024 y diciembre de 2025. Este período corresponde a la aplicación de los cuestionarios, dentro de una investigación cuya planificación había comenzado antes. La recolección en momentos diferentes debe tenerse presente al interpretar las comparaciones, pues no permite asumir que todas las versiones se contestaron bajo condiciones temporales equivalentes. Asimismo, el nivel de confianza empleado en el cálculo inicial no transforma el muestreo voluntario en una muestra probabilística nacional.

3.2. Instrumentos y procedimientos de medición y análisis

3.2.1. Adaptación del cuestionario y formato de aplicación

Las cuatro versiones se elaboraron a partir del Multifactor Leadership Questionnaire de Avolio y Bass (2004). El instrumento sirvió como base temática para presentar afirmaciones reconocibles en un contexto administrativo. Los cuestionarios utilizados fueron adaptaciones: incorporaron paráfrasis, reactivos adicionales y reducciones de contenido. Por ello, sus puntuaciones no se interpretaron como equivalentes automáticos de las escalas originales del MLQ.

La construcción partió de 45 reactivos de base parafraseados, a los que se añadieron cuatro preguntas de atención, una pregunta de carga y 15 formulaciones destinadas a contrastar la consistencia. De ese conjunto de 65 posiciones se derivaron las versiones de menor extensión. Las preguntas complementarias permitieron observar aspectos del proceso de respuesta mientras el participante contestaba sobre estilos y habilidades de liderazgo.

Las respuestas se registraron en una escala de cinco categorías, desde 1, totalmente en desacuerdo, hasta 5, totalmente de acuerdo. Los reactivos se presentaron en cuadrículas y en secciones de aproximadamente diez preguntas. El formulario se implementó en Microsoft Forms, que proporcionó los registros de inicio y finalización empleados para calcular el tiempo total de llenado. No se estableció una condición de respuesta forzada, de modo que se admitieron omisiones y la posibilidad de interrumpir la participación.

La carga percibida se midió con la afirmación «Resolver este cuestionario me resultó muy complicado/tedioso/cansado», incluida al final de las versiones anexadas al estudio. Su puntuación expresó la valoración subjetiva del esfuerzo, en una escala de 1 a 5. Este indicador se trató como predictor de la calidad de respuesta; no se sumó directamente a los siete componentes del índice dependiente.

Los controles de atención incluyeron autoinformes de poca atención y, en la versión de 30 preguntas, una formulación positiva sobre prestar atención. La diferencia de sentido exige considerar la dirección de codificación antes de promediar respuestas. La consistencia se observó mediante pares de afirmaciones de contenido opuesto. La presencia de estas preguntas fue parte del diseño efectivamente aplicado y debe distinguirse de un cuestionario nuevo o de una escala independiente desarrollada para el libro.

3.2.2. Juicio de expertos y prueba piloto

La revisión de contenido estuvo a cargo de cinco expertos con experiencia en administración, investigación cuantitativa y diseño de cuestionarios. Sus perfiles incluyeron trabajo académico y profesional en Ecuador, Perú y Paraguay, además de experiencia en estudios de liderazgo y otros procesos organizacionales. La evaluación atendió a suficiencia, claridad, coherencia y relevancia de los reactivos sometidos a valoración.

El acuerdo de los jueces se resumió mediante la V de Aiken. Los valores informados oscilaron entre 0,917 y 1,000, por encima del criterio de 0,80 utilizado en el estudio con apoyo en Escurra (1988). Las observaciones de los expertos también se

emplearon para mejorar la redacción de algunos enunciados. Este procedimiento aportó evidencia sobre la adecuación del contenido y de su formulación respecto de los propósitos de medición.

Antes del trabajo de campo principal se aplicó un pilotaje a 38 participantes con la versión de 30 preguntas. La encuesta se distribuyó mediante un enlace de Microsoft Forms, siguiendo el procedimiento previsto para la aplicación posterior. El pilotaje no reportó problemas de comprensión y produjo un alfa de Cronbach de 0,68. Este valor se conservó como resultado preliminar, diferenciado de los coeficientes obtenidos después con las muestras de los cuatro grupos.

El juicio de expertos, el pilotaje y el alfa aportaron evidencias distintas. La evaluación de contenido examinó la pertinencia de los reactivos; el pilotaje permitió observar su aplicación; el alfa describió consistencia interna. Ninguno de estos procedimientos, aisladamente, estableció la validez de todas las interpretaciones posibles. En particular, la consistencia interna de las respuestas no sustituyó la evaluación de atención, omisiones y contradicciones que constituía el núcleo de la investigación.

3.2.3. Operacionalización de la calidad de respuesta

La variable dependiente fue un índice de calidad de respuesta construido con siete componentes. Cada componente representó una señal de posible deterioro y se expresó entre 0 y 1. El índice se obtuvo restando la suma de esas señales a un máximo de siete puntos; por ello, una puntuación mayor indicó menos penalizaciones según las reglas de cómputo adoptadas. La literatura sobre calidad en

preguntas de cuadrícula proporcionó antecedentes para seleccionar indicadores complementarios (Gummer et al., 2021; Vehovar et al., 2023).

Tabla 4.

Componentes del índice de calidad de respuesta

Componente	Código	Regla de cómputo
Abandono	c_1 aband	1 si no finalizó el cuestionario; 0 en caso contrario.
Omisión de ítems	c_2 lost	1 si existe al menos un ítem sin respuesta; 0 en caso contrario.
Respuesta lineal	c_3 ddesvest	1 si la desviación estándar de las respuestas es cero.
Extremos y punto medio	c_4 respext	Proporción de respuestas en las categorías 1, 3 o 5.
Desatención reportada	c_5 lackatt_r	Promedio de desatención menos 1, dividido para 4; rango de 0 a 1.
Desatención observada	c_6 lackatt_o	1 si al menos un par contrapuesto recibe el mismo valor de respuesta.
Valores atípicos	c_7 vatip	1 si la distancia de Mahalanobis presenta $p < 0{,}001$.

Fuente: Elaboración propia.

$$CR = 7 - (c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7)$$

El abandono identificó la no finalización del cuestionario entre los registros disponibles. La variable de datos perdidos señaló la presencia de al menos un ítem sin contestar. La respuesta lineal se definió por una desviación estándar igual a cero, mientras que el componente de respuestas extremas y de punto medio correspondió a la proporción de selecciones de las categorías 1, 3 o 5. Esta última regla fue la utilizada por el

estudio; esas categorías pueden expresar opiniones sustantivas y no equivalen por sí solas a respuestas inválidas.

La falta de atención reportada se calculó a partir del promedio de los reactivos de atención, mediante la transformación del promedio menos uno, dividido para cuatro. Su interpretación como desatención requiere que los reactivos estén orientados en la misma dirección. El cuestionario de 30 preguntas contiene enunciados de atención de sentido opuesto, por lo que la recodificación de la formulación positiva es un punto que debe verificarse en la sintaxis de cálculo al reproducir el índice.

La falta de atención observada tomó el valor de uno cuando al menos un par de afirmaciones contrapuestas recibió el mismo valor de respuesta. El indicador registró una señal de inconsistencia conforme a esa regla. La selección del punto medio en ambos enunciados y las diferencias en el número de pares disponibles requieren cautela, porque pueden influir en la probabilidad de activar el indicador. Los valores atípicos se identificaron mediante la distancia de Mahalanobis, con significación inferior a 0,001.

El rango teórico del índice fue de 0 a 7. Para su descripción, el estudio propuso niveles bajo, de 0 a 2,33; medio, de 2,34 a 4,66; y alto, superior a 4,66. Estos cortes organizaron la lectura de las puntuaciones, sin constituir umbrales externos de validez. El alfa de Cronbach se calculó aparte para cada versión: al ser un resultado del grupo y no una característica individual variable, no se incorporó como uno de los siete sumandos del índice.

3.2.4. Recolección de datos y condiciones de participación

La aplicación se realizó mediante enlaces a los formularios digitales. Las instrucciones solicitaron valorar afirmaciones sobre estilos y habilidades de liderazgo de trabajadores de Ecuador. La respuesta fue autoadministrada y voluntaria. El diseño no exigió completar todos los reactivos, lo que permitió observar omisiones como parte de la información de interés, en lugar de impedirlas mediante una restricción de la plataforma.

La finalidad metodológica de evaluar la atención y la consistencia no se explicitó en la presentación del cuestionario. El propósito declarado al participante se centró en el contenido de liderazgo. Esta decisión buscó reducir cambios de conducta provocados por conocer que se observaba la calidad de las respuestas. La descripción del procedimiento debe reconocer ese encuadre y distinguirlo de una evaluación organizacional destinada a calificar a personas o instituciones.

Desde el punto de vista ético, el tratamiento de esta información exige proteger la identidad de los participantes y evitar atribuciones individuales de desempeño a partir de indicadores metodológicos. La voluntariedad y la posibilidad de no continuar fueron condiciones del procedimiento descrito. La comunicación de los resultados se realizó en términos agregados por versión, sin presentar evaluaciones personales de liderazgo ni identificar trabajadores a partir de patrones de respuesta.

La diferencia entre las marcas de inicio y finalización se convirtió a minutos. Ese registro representó tiempo transcurrido y pudo incluir pausas o interrupciones. No se contó con una medición directa del tiempo de atención activa ni con observaciones de las actividades realizadas durante las

pausas. De igual manera, los indicadores de abandono y omisión describieron los registros recuperados para el análisis; no permiten conocer, por sí solos, cuántas personas recibieron el enlace y nunca iniciaron o enviaron una respuesta.

3.2.5. Procesamiento y análisis estadístico

El procesamiento comenzó en Excel mediante la codificación y etiquetado de variables, la integración de las cuatro bases y el cálculo de los indicadores. Se generó la duración a partir de las marcas temporales y se construyeron los componentes del índice de calidad. Posteriormente, se trabajó en SPSS para calcular la distancia de Mahalanobis y efectuar los análisis estadísticos. La revisión de la base buscó identificar problemas de registro sin eliminar automáticamente las señales de baja calidad que constituían el resultado de interés.

El análisis descriptivo incluyó tamaños de muestra, promedios de tiempo, edad y carga, porcentajes de los indicadores y medias del índice de calidad. Se calculó el alfa de Cronbach para cada versión y se aplicó la prueba de Kolmogorov-Smirnov a calidad, tiempo y edad. La comparación global se realizó mediante un análisis de varianza de un factor, seguido de pruebas t para muestras independientes entre pares de versiones.

El análisis dentro de cada grupo empleó una regresión lineal en los coeficientes, con tiempo lineal y tiempo al cuadrado, carga percibida, una variable indicadora de género femenino y edad. La siguiente expresión reproduce la especificación utilizada, sin término constante.

$$RQ_i = \beta_1 \cdot t_i^2 + \beta_2 \cdot t_i + \beta_3 \cdot ca_i + \beta_4 \cdot df_i + \beta_5 \cdot ed_i + \varepsilon_i$$

En esta expresión, RQ representa la calidad de respuesta; t, el tiempo en minutos; ca, la carga percibida; df, la variable indicadora que tomó el valor de uno para género femenino; ed, la edad en años; y ε, el término de error. La exclusión de la constante respondió a la intención de vincular ausencia de respuesta con ausencia de calidad. Sin embargo, en una regresión multivariable, fijar el tiempo en cero no anula los términos de edad, carga o género. La especificación se interpreta, por tanto, como una decisión del modelo y no como una garantía matemática de calidad nula cuando el tiempo es cero.

La interpretación consideró la no normalidad, los tamaños desiguales y la falta de diagnósticos completos de varianzas y residuos. Las comparaciones por pares conservaron las significaciones informadas, sin atribuirles un ajuste por multiplicidad no documentado. Los resultados corresponden al análisis realizado y deben distinguirse de las verificaciones adicionales que requieren la base y la sintaxis originales.

CAPÍTULO IV

RESULTADOS DE LA INVESTIGACIÓN SOBRE CALIDAD DE RESPUESTA

La comparación de las cuatro versiones mostró diferencias en la calidad media de respuesta. Los cuestionarios de 30 y 40 preguntas obtuvieron medias superiores a las observadas en las versiones de 50 y 65 preguntas. Al mismo tiempo, el comportamiento de los indicadores no fue uniforme: algunas señales aumentaron con la extensión y otras presentaron variaciones sin una secuencia estrictamente creciente. Esta combinación de resultados permite analizar tanto la puntuación conjunta como sus componentes.

El capítulo presenta la muestra, la consistencia interna, los resultados descriptivos y los contrastes estadísticos. Los valores corresponden a los análisis de la investigación. Las comparaciones aritméticas que se utilizan para explicar diferencias de medias o porcentajes no sustituyen una nueva estimación con datos individuales. La interpretación de los mecanismos y el contraste con estudios previos se desarrollan en el capítulo siguiente.

4.1. Características de los grupos e indicadores descriptivos

4.1.1. Distribución de los participantes por versión

La muestra total estuvo integrada por 473 participantes. El grupo de 30 preguntas reunió 186 personas, equivalentes al 39,32 % del total. La versión de 40 preguntas fue respondida por 92 participantes, el 19,45 %; la de 50 preguntas, por 84, el 17,76 %; y la de 65 preguntas, por 111, el 23,47 %. Los porcentajes se calcularon utilizando los 473 registros como denominador.

Tabla 5.

Participantes por versión del cuestionario

Preguntas	Participantes	Porcentaje
30	186	39,32 %
40	92	19,45 %
50	84	17,76 %
65	111	23,47 %
Total	473	100,00 %

Fuente: Elaboración propia.

Todos los grupos superaron el mínimo operativo de 80 participantes. El grupo de referencia fue el más numeroso y el de 50 preguntas presentó el menor tamaño. Esta distribución permitió desarrollar las comparaciones previstas, aunque las estimaciones de los grupos no contaron con idéntica precisión. La desigualdad muestral también hace relevante considerar la variabilidad de las puntuaciones al valorar los resultados del análisis de varianza.

El total de participantes constituye el conjunto utilizado para los análisis del estudio. No representa una tasa de respuesta, porque ese indicador necesitaría un denominador de personas invitadas o elegibles. Tampoco debe interpretarse como una muestra proporcional de los sectores público y privado del país. Su utilidad radica en la comparación de las cuatro condiciones de cuestionario dentro del reclutamiento realizado.

4.1.2. Consistencia interna de las cuatro versiones

Los coeficientes alfa de Cronbach fueron 0,836 para la versión de 30 preguntas, 0,825 para la de 40, 0,916 para la de 50 y 0,945 para la de 65. Los cuatro valores superaron 0,80 y fueron mayores que el coeficiente preliminar de 0,68 obtenido en el pilotaje de la versión de referencia. Los resultados mostraron consistencia interna en las respuestas analizadas dentro de cada grupo.

Tabla 6.

Consistencia interna de las versiones

Preguntas	Participantes	Alfa de Cronbach
30	186	0,836
40	92	0,825
50	84	0,916
65	111	0,945

Fuente: Elaboración propia.

La comparación descriptiva situó los valores más elevados de alfa en las versiones de 50 y 65 preguntas. Este resultado no coincidió con el orden de las medias del índice de calidad, que fueron mayores en las versiones de 30 y 40. La divergencia es relevante porque muestra que una mayor consistencia interna del conjunto de respuestas puede coexistir con más señales de desatención, omisión o contradicción.

El alfa depende de las relaciones entre los ítems y de la cantidad de reactivos que se incluyen. Por ello, el aumento observado en las versiones largas no permite concluir, de forma aislada, que sus participantes respondieran con mayor cuidado. Los instrumentos también variaron en contenido y número de controles. En esta investigación, el alfa

complementó la descripción de la medición, mientras que el índice de calidad resumió señales definidas a nivel individual.

4.1.3. Tiempo de llenado y características de los grupos

El tiempo promedio de llenado aumentó con la extensión de las versiones. Fue de 7,03 minutos en el cuestionario de 30 preguntas, 7,40 en el de 40, 8,79 en el de 50 y 13,56 en el de 65. Entre las versiones extremas se observó una diferencia de 6,53 minutos. El aumento más marcado se produjo al pasar de la condición de 50 a la de 65 preguntas, con 4,77 minutos adicionales en promedio.

Tabla 7.

Tiempo de llenado edad y carga percibida

Indicador	30	40	50	65
Tiempo medio en minutos	7,03	7,40	8,79	13,56
Edad media en años	33,00	35,00	30,75	34,19
Carga media de 1 a 5	1,93	2,27	2,73	2,77

Fuente: Elaboración propia.

La edad media fue de 33 años en el grupo de referencia, 35 en la versión de 40 preguntas, 30,75 en la de 50 y 34,19 en la de 65. Por tanto, los grupos no tuvieron exactamente la misma composición etaria. La edad se incorporó después como predictor en las regresiones, aunque estos modelos se estimaron por separado y no equivalen a una comparación global de versiones ajustada por todas las diferencias entre participantes.

La carga percibida promedio pasó de 1,93 a 2,27, 2,73 y 2,77, respectivamente. La secuencia fue compatible con una mayor valoración subjetiva del esfuerzo en las versiones más extensas. Sin embargo, la cercanía de las medias de 50 y 65 preguntas

contrastó con el aumento de su duración. Este resultado descriptivo muestra que el tiempo transcurrido y la experiencia subjetiva de carga aportaron información relacionada, pero no intercambiable.

Las cifras de duración son promedios de lo ocurrido durante el llenado. No constituyen tiempos recomendados ni puntos de máxima calidad. Una media de 7,03 minutos en el grupo de referencia informa sobre su experiencia de aplicación, mientras que determinar una duración óptima exige un análisis específico de la curva tiempo-calidad y de la incertidumbre asociada a sus parámetros.

4.1.4. Comportamiento de los componentes de calidad

La falta de atención observada presentó una secuencia creciente: 57,53 % en la versión de 30 preguntas, 70,65 % en la de 40, 82,14 % en la de 50 y 87,39 % en la de 65. Este indicador correspondió a la detección de al menos un par de afirmaciones contrapuestas con igual respuesta. La diferencia entre los grupos de 30 y 65 preguntas fue de 29,86 puntos porcentuales, conforme a la regla operacional utilizada.

La falta de atención reportada fue de 13,44 %, 19,97 %, 32,61 % y 28,62 %. Estas cifras expresan promedios del indicador normalizado multiplicados por cien, no porcentajes de personas clasificadas mediante un punto de corte. El mayor valor se observó en la versión de 50 preguntas. Por consiguiente, el autoinforme de desatención no siguió exactamente la misma secuencia que el indicador basado en contradicciones.

Las respuestas lineales fueron más frecuentes en las versiones de 50 y 65 preguntas, con 7,14 % y 7,21 %, frente a

2,16 % y 3,26 % en las de 30 y 40. La presencia de al menos un ítem omitido alcanzó 15,59 %, 11,96 %, 21,43 % y 18,01 %. El abandono registrado fue de 1,08 %, 3,26 %, 7,14 % y 4,50 %. En estos dos últimos indicadores, el grupo de 50 preguntas presentó los porcentajes más altos.

Tabla 8.

Indicadores de calidad por versión

Indicador en porcentaje	30	40	50	65
Valores atípicos	17,20	10,87	21,40	15,30
Desatención reportada	13,44	19,97	32,61	28,62
Desatención observada	57,53	70,65	82,14	87,39
Extremos y punto medio	72,58	71,22	70,19	71,55
Respuestas lineales	2,16	3,26	7,14	7,21
Personas con ítems omitidos	15,59	11,96	21,43	18,01
Abandono registrado	1,08	3,26	7,14	4,50

Fuente: Elaboración propia.

La proporción media de respuestas en las categorías 1, 3 y 5 permaneció relativamente próxima entre las versiones, con valores de 70,19 % a 72,58 %. Los valores atípicos oscilaron entre 10,87 % y 21,40 %, sin aumentar de manera continua con la longitud. Estos resultados muestran que las diferencias del índice conjunto no se debieron a un empeoramiento uniforme de todas las dimensiones. La lectura por componentes permitió identificar qué señales variaron más y cuáles permanecieron relativamente estables.

4.1.5. Calidad media y distribución de las variables

Las medias del índice de calidad fueron 5,1935 para 30 preguntas, 5,0880 para 40, 4,5791 para 50 y 4,6740 para 65. Las versiones breves ocuparon las primeras posiciones descriptivas. La versión de 65 preguntas obtuvo una media ligeramente superior a la de 50, por lo que el patrón global no consistió en una disminución estricta con cada incremento de longitud.

Tabla 9.
Calidad media de respuesta por versión

Preguntas	Media del índice	Nivel de la media
30	5,1935	Alto
40	5,0880	Alto
50	4,5791	Medio
65	4,6740	Alto

Fuente: Elaboración propia.

La clasificación descriptiva propuesta en el estudio situó las medias de 30, 40 y 65 preguntas por encima de 4,66 y la de 50 dentro del intervalo medio. Esta ubicación de las medias no permite conocer cuántas personas pertenecieron a cada nivel. En particular, el promedio de la versión de 65 preguntas quedó muy próximo al corte, de modo que la etiqueta de nivel alto no elimina la diferencia que mantuvo respecto de las versiones breves.

La prueba de Kolmogorov-Smirnov rechazó la hipótesis de normalidad del índice de calidad en los cuatro grupos. Los valores de significación fueron inferiores a 0,001 para 30 preguntas, 0,003 para 40, 0,012 para 50 e inferiores a 0,001 para 65. En el tiempo de llenado, las cuatro significaciones fueron

inferiores a 0,001. La edad presentó p = 0,005, p = 0,078, p = 0,012 y p = 0,001, respectivamente.

Tabla 10.
Significación de la prueba de Kolmogorov Smirnov

Variable	30	40	50	65
Calidad	< 0,001	0,003	0,012	< 0,001
Tiempo	< 0,001	< 0,001	< 0,001	< 0,001
Edad	0,005	0,078	0,012	0,001

Fuente: Elaboración propia.

Estos resultados describieron distribuciones que se apartaron de la normalidad en la mayoría de los casos evaluados. Su importancia analítica reside en que los contrastes posteriores deben valorarse junto con las condiciones de aplicación de las pruebas paramétricas. El rechazo de normalidad no convierte por sí mismo en inutilizable la información, pero requiere considerar variabilidad, tamaños de grupo y diagnósticos del modelo antes de atribuir una precisión definitiva a las estimaciones.

4.2. Comparaciones de calidad y asociaciones dentro de cada grupo

4.2.1. Comparación global entre las cuatro versiones

El análisis de varianza identificó diferencias globales en la calidad media de respuesta. El estadístico fue F(3, 469) = 8,712, con p < 0,001. En consecuencia, se rechazó la hipótesis nula de igualdad de las cuatro medias bajo el modelo utilizado. El resultado indicó que al menos una media difería de las

restantes, sin establecer todavía cuáles pares de versiones explicaban esa diferencia.

Tabla 11.
Análisis de varianza de la calidad de respuesta

Fuente	Suma de cuadrados	gl	Media cuadrática
Entre grupos	32,767	3	10,922
Dentro de grupos	587,968	469	1,254
Total	620,735	472	—

Fuente: Elaboración propia.

La suma de cuadrados entre grupos fue de 32,767 y la correspondiente a la variación dentro de los grupos fue de 587,968. La suma total alcanzó 620,735. Estos valores muestran que las diferencias entre versiones coexistieron con una variabilidad considerable entre las personas que respondieron una misma versión. Por tanto, la extensión no agotó la explicación del comportamiento del índice.

La significación global justificó examinar las comparaciones por pares. Su interpretación quedó vinculada al diseño de aplicación y a los supuestos del procedimiento estadístico. Aunque la extensión se modificó deliberadamente, el contraste no aisló por sí mismo las diferencias de composición o de período de recolección. El resultado principal fue una diferencia estadística entre condiciones observadas, cuya explicación se discutió a la luz de esos elementos.

4.2.2. Comparaciones entre pares de cuestionarios

La comparación entre 30 y 40 preguntas produjo p = 0,437. La diferencia descriptiva entre sus medias fue de 0,1055 puntos a favor de la versión de 30 preguntas. No se detectó una diferencia estadísticamente significativa en ese contraste. Este resultado no equivale a demostrar que ambas versiones tengan exactamente la misma calidad, pues no se aplicó una prueba de equivalencia con márgenes previamente definidos.

Tabla 12.

Comparaciones nominales entre pares de versiones

Preguntas comparadas	Diferencia de medias	Valor de p
30 frente a 40	0,1055	0,437
30 frente a 50	0,6144	< 0,001
30 frente a 65	0,5195	< 0,001
40 frente a 50	0,5089	0,005
40 frente a 65	0,4140	0,009
50 frente a 65	−0,0949	0,590

Fuente: Elaboración propia.

Las versiones de 30 y 40 preguntas presentaron medias mayores que la versión de 50. Las significaciones fueron p < 0,001 y p = 0,005, respectivamente. También superaron a la versión de 65 preguntas, con p < 0,001 para la comparación de 30 frente a 65 y p = 0,009 para 40 frente a 65. Las diferencias descriptivas oscilaron entre 0,4140 y 0,6144 puntos en estos cuatro contrastes.

Entre 50 y 65 preguntas no se detectó diferencia significativa, con p = 0,590. La media de la versión de 65 fue 0,0949 puntos mayor. El patrón conjunto distinguió así las dos

versiones breves de las dos extensas en las comparaciones informadas, sin establecer una gradación de pérdida de calidad proporcional a cada pregunta añadida.

Los valores de p se presentan como significaciones nominales de las pruebas t realizadas. El análisis no documentó un ajuste por las seis comparaciones simultáneas. Esta condición importa especialmente para interpretar resultados próximos al criterio de decisión. La evidencia sostiene la conveniencia de examinar la extensión en las condiciones del estudio, pero no define un límite exacto entre 41 y 49 preguntas, longitudes que no fueron evaluadas.

4.2.3. Asociación entre tiempo y calidad de respuesta

Las regresiones estimadas para las versiones de 30, 40 y 50 preguntas presentaron coeficientes positivos para el tiempo y negativos para el tiempo al cuadrado. En la versión de 30 preguntas fueron 0,092 y −0,001, ambos con p = 0,006. En la de 40, los valores fueron 0,149, con p = 0,006, y −0,002, con p = 0,011. En la de 50 se obtuvieron 0,128, con p = 0,001, y −0,001, con p = 0,003.

La combinación de un término lineal positivo y otro cuadrático negativo fue compatible con una curva cóncava: dentro de la especificación del modelo, la calidad aumentó con el tiempo hasta un punto de máximo y después disminuyó. Esta forma estadística sugiere que interpretar toda duración adicional como una mejora sería inadecuado. También indica que el tiempo aporta información cuyo sentido depende del tramo de la curva y de las demás variables incluidas.

En el grupo de 65 preguntas, el término lineal fue 0,054, con p = 0,003, y el término cuadrático apareció redondeado a 0,000,

con p = 0,005. El redondeo impide recuperar su signo y magnitud exactos. Por ello, no corresponde atribuir a esta versión la misma curvatura con igual certeza ni calcular su punto de máximo a partir del valor mostrado.

Tabla 13.

Coeficientes de regresión por versión

Preguntas	Predictor	Coeficiente	Valor de p
30	Tiempo al cuadrado	−0,001	0,006
30	Tiempo	0,092	0,006
30	Carga	0,379	< 0,001
30	Género femenino	0,878	< 0,001
30	Edad	0,104	< 0,001
40	Tiempo al cuadrado	−0,002	0,011
40	Tiempo	0,149	0,006
40	Carga	0,291	0,009
40	Género femenino	0,882	0,010
40	Edad	0,079	< 0,001
50	Tiempo al cuadrado	−0,001	0,003
50	Tiempo	0,128	0,001
50	Carga	0,320	0,003
50	Género femenino	0,088	0,827
50	Edad	0,090	< 0,001
65	Tiempo al cuadrado	0,000	0,005
65	Tiempo	0,054	0,003
65	Carga	0,367	< 0,001
65	Género femenino	1,101	0,001

Preguntas	Predictor	Coeficiente	Valor de p
65	Edad	0,070	< 0,001

Fuente: Elaboración propia.

Los coeficientes publicados tampoco permiten establecer tiempos óptimos suficientemente precisos para las otras versiones. Dividir el coeficiente lineal por el doble del cuadrático redondeado puede producir valores sensibles a pequeñas variaciones no visibles en la tabla. Además, se requieren el rango observado, la incertidumbre de la estimación y la evaluación del ajuste. El resultado empírico defendible es la presencia de una relación no lineal en las tres primeras versiones, sin convertirla en una recomendación universal de minutos de respuesta.

4.2.4. Asociación de la carga percibida y la edad

La carga percibida presentó coeficientes positivos en las cuatro regresiones. Los valores fueron 0,379 para 30 preguntas, 0,291 para 40, 0,320 para 50 y 0,367 para 65. Sus significaciones fueron $p < 0,001$, $p = 0,009$, $p = 0,003$ y $p < 0,001$, respectivamente. En los modelos estimados, mayores puntuaciones de carga se asociaron con mayores puntuaciones de calidad, manteniendo constantes las restantes variables incluidas.

Este resultado se refiere a una asociación ajustada dentro de cada grupo. Es diferente de la comparación descriptiva entre versiones, en la que las formas largas mostraron mayor carga media y menor calidad media que las breves. Ambos resultados pueden coexistir porque responden a preguntas distintas: uno compara condiciones de cuestionario y el otro compara participantes dentro de una misma condición.

La edad también presentó coeficientes positivos y significativos en las cuatro versiones: 0,104, 0,079, 0,090 y 0,070, todos con $p < 0,001$. Las estimaciones sugirieron una relación positiva entre edad y calidad dentro de las muestras estudiadas y bajo la especificación utilizada. No se estimó un término cuadrático de edad, por lo que los resultados no demuestran una curva de aumento y descenso a lo largo de la vida.

La interpretación de estos coeficientes debe conservar su carácter condicional. La relación positiva de la carga no prueba que hacer más difícil un cuestionario mejore las respuestas; la relación con edad no garantiza mayor calidad para cualquier persona de más años. Motivación, experiencia laboral y otras características podrían intervenir, pero no fueron medidas mediante variables específicas que permitieran contrastar esos mecanismos.

4.2.5. Asociación del género y síntesis de los contrastes

La variable indicadora de género femenino presentó coeficientes de 0,878 en la versión de 30 preguntas, 0,882 en la de 40 y 1,101 en la de 65. Las significaciones fueron $p < 0,001$, $p = 0,010$ y $p = 0,001$, respectivamente. En esas tres regresiones se observó una asociación positiva con el índice de calidad respecto de la categoría codificada como cero, una vez consideradas las otras variables del modelo.

En la versión de 50 preguntas, el coeficiente fue 0,088 y la significación $p = 0,827$. No se detectó una asociación estadísticamente significativa en ese grupo. La diferencia entre resultados significativos y no significativos tampoco demuestra por sí sola que el efecto del género cambie entre versiones; para evaluar ese cambio habría sido necesario

contrastar directamente una interacción o una diferencia entre coeficientes.

La lectura conjunta de los análisis identificó cuatro resultados: diferencias de calidad entre las versiones breves y extensas en los contrastes nominales; una relación temporal cóncava en las versiones de 30, 40 y 50 preguntas; asociaciones positivas de carga y edad en los cuatro grupos; y una asociación de género que no alcanzó significación en la versión de 50 preguntas. Cada resultado corresponde a una comparación concreta y conserva el alcance del modelo que lo produjo.

Estos hallazgos respondieron al problema de investigación sin extenderlo hacia la creación de un instrumento administrativo diferente. Su aporte consiste en mostrar cómo cambiaron las señales de calidad en las versiones aplicadas y qué variables se relacionaron con el índice. La discusión siguiente examina su significado, su relación con otros estudios y las recomendaciones que pueden formularse dentro de ese alcance.

CAPÍTULO V

DISCUSIÓN DE LOS HALLAZGOS Y CONCLUSIONES DE LA INVESTIGACIÓN

La investigación aportó evidencia de una relación entre la extensión del cuestionario y la calidad de respuesta en las condiciones estudiadas. Las versiones de 30 y 40 preguntas obtuvieron medias más altas que las de 50 y 65, mientras que los indicadores mostraron comportamientos diferentes entre sí. El resultado principal debe comprenderse junto con el contenido del instrumento, la forma de construir el índice y el reclutamiento de los participantes.

La discusión examina esas relaciones y las contrasta con investigaciones sobre longitud, carga y comportamiento de respuesta. A partir de los hallazgos se formulan conclusiones y recomendaciones para estudios administrativos que utilicen encuestas digitales. Las recomendaciones se derivan de la investigación realizada y mantienen como objeto la calidad de respuesta; no suponen la elaboración ni la validación de un nuevo cuestionario de gestión.

5.1. Interpretación de los resultados y contraste con la literatura

5.1.1. Extensión del cuestionario y diferencias de calidad

La mayor calidad media de las versiones de 30 y 40 preguntas sugiere que ampliar el formulario puede tener un costo metodológico. En esta investigación, ese costo se observó en una combinación de señales de desatención, inconsistencia y otras penalizaciones del índice. La diferencia entre 30 y 40 preguntas fue pequeña y no significativa, mientras que las comparaciones de ambas versiones con las de 50 y 65 alcanzaron significación nominal. El patrón identifica una zona de contraste entre las longitudes evaluadas, aunque no fija el punto exacto en que comienza el deterioro.

La dirección general es compatible con Eisele et al. (2022), quienes compararon cuestionarios de 30 y 60 ítems administrados repetidamente a estudiantes y observaron mayor carga y deterioro de indicadores de cantidad y calidad de datos con cuestionarios más largos. Su diseño de muestreo de experiencias difiere de una encuesta administrativa aplicada a trabajadores. Por ello, la coincidencia respalda la relevancia de estudiar la longitud, sin trasladar automáticamente sus magnitudes o condiciones de aplicación al presente estudio.

El efecto de la extensión tampoco es uniforme en toda la literatura. Cernat et al. (2024) encontraron una calidad de medición relativamente similar al comparar diseños de encuesta y versiones largas y breves en la muestra alemana del European Values Study. Sus análisis incluyeron indicadores y pruebas de equivalencia de escalas diferentes del índice utilizado aquí. La divergencia indica que las conclusiones dependen de qué se entiende por calidad, cómo se modifica el cuestionario y qué población lo responde.

Desde el punto de vista explicativo, una mayor extensión podría favorecer estrategias de simplificación si el esfuerzo acumulado supera la disposición del participante a seguir evaluando cada enunciado. Sin embargo, el estudio no midió directamente el agotamiento cognitivo ni manipuló la motivación. La fatiga constituye una explicación plausible del patrón, no un mecanismo causal demostrado. También deben considerarse los cambios de contenido y del número de controles entre versiones.

La recomendación que se desprende de los datos es evaluar con especial cuidado el paso de versiones de 30 o 40 preguntas a versiones de 50 o 65 en instrumentos semejantes. No resulta

justificable convertir 40 preguntas en una frontera universal de validez. Tampoco corresponde invalidar investigaciones anteriores únicamente por superar esa longitud: su calidad depende de sus propios datos, indicadores, participantes y condiciones de aplicación.

5.1.2. Atención y consistencia como dimensiones de la calidad

La falta de atención observada fue el componente que mostró la secuencia creciente más clara entre las cuatro versiones. El aumento de 57,53 % a 87,39 % sugiere que las contradicciones detectadas por la regla del estudio adquirieron mayor presencia en los instrumentos extensos. No obstante, la variable registró al menos una coincidencia de respuesta en pares contrapuestos. Al aumentar el número de pares evaluables, también pueden aumentar las oportunidades de activar esa señal, incluso sin un cambio proporcional en la conducta de cada participante.

Esta característica exige interpretar el indicador junto con otros resultados. Las respuestas lineales fueron más frecuentes en las versiones largas, lo que aporta una señal adicional compatible con menor diferenciación entre reactivos. En cambio, las respuestas extremas y de punto medio variaron poco. La omisión y el abandono alcanzaron sus máximos en la versión de 50 preguntas y descendieron en la de 65. El deterioro de calidad, por tanto, no adoptó una forma idéntica en todas las dimensiones.

Andreadis y Kartsounidou (2020) también observaron resultados diferentes según el indicador al dividir un cuestionario largo. Algunas medidas mejoraron, pero no

encontraron diferencias significativas en omisión de ítems, respuestas rápidas o respuestas lineales; además, la participación acumulada en las dos partes fue inferior a la del cuestionario completo. Su trabajo respalda una lectura por dimensiones y evita suponer que reducir o dividir un formulario resuelve de manera uniforme todos los problemas de calidad.

Los coeficientes alfa del presente estudio refuerzan esa necesidad de evaluación conjunta. Las versiones de 50 y 65 preguntas alcanzaron los valores más altos de consistencia interna, aunque obtuvieron menores medias de calidad que las breves. Una estructura de respuestas internamente relacionada puede coexistir con patrones repetitivos o con fallas de atención. La decisión sobre el uso de una base no debería descansar exclusivamente en que su alfa supere un umbral habitual.

La contribución metodológica consiste en hacer visibles varias señales que pueden quedar ocultas tras una puntuación global. A su vez, el índice debe interpretarse conforme a sus reglas. Elegir extremos o el punto medio puede ser legítimo, y una coincidencia en un par contrapuesto requiere contexto. Las penalizaciones aportan indicadores de revisión; no autorizan a calificar como falsa cada respuesta ni a atribuir falta de honestidad a una persona.

5.1.3. Tiempo de respuesta y relación no lineal

Los términos temporal lineal y cuadrático de las versiones de 30, 40 y 50 preguntas fueron compatibles con una relación cóncava entre duración y calidad. La interpretación sustantiva es que dedicar más tiempo puede asociarse con un mejor

procesamiento hasta cierto tramo, mientras que una duración adicional no mantiene indefinidamente esa relación. Este patrón ofrece una explicación más específica que asumir una correspondencia siempre positiva o siempre negativa entre rapidez y calidad.

El antecedente de Izaguirre Olmedo et al. (2025) examinó una encuesta de liderazgo adaptada del MLQ en 224 participantes y también identificó un comportamiento temporal no lineal. Ese trabajo aporta continuidad a la pregunta metodológica, pero su muestra y su cuestionario no son intercambiables con los cuatro grupos aquí analizados. Una duración propuesta en aquel antecedente no puede utilizarse como tiempo óptimo de las versiones de 30, 40, 50 y 65 preguntas.

Un tiempo muy breve puede reflejar lectura superficial, aunque también familiaridad con los contenidos o facilidad para decidir. Un tiempo largo puede incluir mayor deliberación o pausas ajenas a la encuesta. La medición empleada registró tiempo transcurrido entre inicio y finalización. Por ello, no permite separar esas situaciones ni identificar por sí sola si una persona realizó otras tareas durante el llenado.

La multitarea constituye una línea explicativa pertinente en encuestas en línea, estudiada por Décieux (2024). En la presente investigación, sin embargo, no se registraron directamente cambios de actividad, llamadas o interrupciones laborales. Atribuir la caída de calidad exclusivamente a esos procesos excedería la evidencia. El resultado invita a medirlos en nuevas aplicaciones para evaluar si explican parte de la asociación temporal.

El estudio también muestra por qué debe distinguirse la forma de una curva de la precisión de su máximo. Los coeficientes cuadráticos se publicaron con pocos decimales y el correspondiente a 65 preguntas quedó redondeado a cero. En estas condiciones, no se establece una recomendación concreta de minutos. La duración conserva valor como indicador de seguimiento y análisis, pero requiere información adicional para convertirse en un criterio de diseño o exclusión de registros.

5.1.4. Carga percibida y esfuerzo de respuesta

La asociación positiva entre carga percibida y calidad fue uno de los resultados que exigió mayor atención interpretativa. En las cuatro regresiones, quienes informaron una mayor dificultad tendieron a obtener puntuaciones más altas de calidad, considerando las demás variables del modelo. Al mismo tiempo, las versiones largas tuvieron mayor carga promedio y menor calidad media que las breves. El contraste muestra que una relación entre grupos no debe confundirse con la relación entre personas dentro de cada grupo.

El hallazgo difiere de Kunz y Gummer (2025), quienes encontraron que una mayor carga percibida se relacionaba con peor calidad en varios indicadores de una encuesta web. Su estudio distinguió la carga objetiva de la subjetiva y realizó mediciones de percepción en diferentes momentos. En esta investigación, la carga se recogió mediante una afirmación final que reunió dificultad, tedio y cansancio. Las diferencias de medición y diseño limitan una comparación directa de los coeficientes.

Una interpretación posible es que algunos participantes identificaran como carga el esfuerzo que invirtieron al responder cuidadosamente. Otra posibilidad es que el reactivo reuniera experiencias distintas: un formulario puede parecer complicado por su contenido, tedioso por repetición o cansado por la situación de aplicación. El análisis no separó esas dimensiones. En consecuencia, la explicación mediante mayor involucramiento cognitivo debe mantenerse como hipótesis y no como una característica confirmada de los participantes.

La implicación práctica es conservar la medición de carga como información complementaria y examinar su relación con señales observables. Los resultados no respaldan aumentar intencionalmente la dificultad o la repetición para obtener mejores respuestas. Más bien, justifican revisar qué experiencia expresa el participante al valorar la carga y si la formulación utilizada permite distinguir esfuerzo productivo de molestias que dificultan continuar.

5.1.5. Edad y género en el comportamiento de respuesta

La edad se asoció positivamente con la calidad en las cuatro versiones. Este resultado se obtuvo en trabajadores que participaron voluntariamente en una encuesta digital. La experiencia con situaciones laborales podría facilitar la comprensión de afirmaciones sobre liderazgo, pero esa explicación no fue contrastada mediante una medición específica de experiencia o trayectoria. La asociación observada debe describirse dentro de la muestra y del contenido del cuestionario.

El trabajo de Schneider et al. (2024) examinó funcionamiento cognitivo y calidad de respuesta en estudios de envejecimiento.

Su objeto permite reconocer que la edad cronológica y las capacidades relevantes para contestar no son equivalentes. Comparar sus resultados con los presentes exige atender a las poblaciones y a las variables medidas. La asociación positiva encontrada no contradice toda evidencia sobre dificultades de respuesta en edades avanzadas ni demuestra que la calidad aumente de forma ilimitada con la edad.

La variable indicadora de género femenino se asoció positivamente con calidad en tres versiones, pero no alcanzó significación en la de 50 preguntas. Este resultado describe una relación estadística de las muestras observadas. No establece una diferencia inherente de capacidad, atención o compromiso entre géneros, y tampoco justifica seleccionar o excluir participantes por esa característica.

Garbarski et al. (2025) estudiaron cómo los formatos de respuesta para preguntar el género afectan su medición y la calidad de datos. Ese antecedente es relevante para la forma de operacionalizar la variable, pero no demuestra por sí mismo que las mujeres produzcan respuestas de mayor calidad. En el presente estudio, explicar la asociación encontrada requeriría evaluar composición de los grupos, experiencia, motivación y otras variables relacionadas, además de contrastar directamente si las diferencias cambian según la extensión.

5.2. Alcance de los hallazgos y aportes para la investigación administrativa

5.2.1. Implicaciones para encuestas digitales administrativas

El aporte aplicado del estudio es mostrar que la extensión debe justificarse como parte de la calidad de medición. En las

versiones analizadas, añadir preguntas no se tradujo en mejores puntuaciones del índice. Quien diseña una encuesta administrativa debe valorar la información que aporta cada reactivo frente al esfuerzo de respuesta que introduce. Esa valoración resulta especialmente pertinente cuando un formulario crece por incorporación sucesiva de temas o solicitudes de distintas áreas.

La evidencia favorece considerar versiones de 30 a 40 preguntas como referencia inicial para instrumentos similares, siempre que mantengan el contenido necesario. Esta orientación es contextual: el estudio no evaluó cuestionarios de menos de 30 preguntas ni todas las longitudes entre las condiciones aplicadas. Reducir indiscriminadamente un instrumento puede eliminar dimensiones esenciales y afectar la interpretación de las puntuaciones, aun cuando disminuya el tiempo de respuesta.

Los resultados también respaldan registrar duración y conservar indicadores de atención, omisión y consistencia. El uso conjunto de esas señales permite comprender mejor qué ocurre durante la aplicación. En particular, un alfa elevado no debería emplearse como argumento suficiente para omitir la revisión de otros patrones. La utilidad de una base para decisiones administrativas depende de que las inferencias estén apoyadas tanto en el contenido de las preguntas como en la calidad del proceso de respuesta.

La aplicación de estas orientaciones requiere distinguir evaluación metodológica de evaluación del personal. Un indicador de inconsistencia no es una medida de desempeño laboral. Las organizaciones deben utilizarlo para revisar el instrumento, las condiciones de recolección y la solidez del

análisis, evitando convertir una señal estadística en un juicio individual sobre la persona que colaboró con la investigación.

5.2.2. Aportes de la medición conjunta y de la comparación de versiones

La comparación de cuatro versiones permitió estudiar una misma pregunta metodológica desde dos niveles. Entre grupos se examinó la diferencia de calidad asociada con condiciones de longitud. Dentro de cada grupo se exploraron relaciones entre calidad, tiempo, carga y características personales. La combinación amplió la descripción del fenómeno, aunque los modelos intragrupo no reemplazaron un contraste global ajustado de la extensión.

La construcción de un índice con siete componentes facilitó la síntesis, pero también mostró la necesidad de explicar su composición. Un valor numérico adquiere sentido cuando se conoce qué conductas lo reducen, cómo se ponderan y qué registros lo sustentan. En este caso, los componentes recibieron el mismo peso máximo, pese a representar señales distintas. Esa decisión permitió el cálculo operativo y, al mismo tiempo, condicionó el significado de una diferencia de puntuación.

La lectura por componentes evitó atribuir el resultado exclusivamente al abandono o a la omisión. La señal de contradicción varió de manera más marcada, mientras que la selección de extremos y del punto medio permaneció relativamente estable. El índice debe acompañarse de ese desglose para que el lector pueda identificar la contribución de cada dimensión y valorar si los resultados serían sensibles a otras reglas de puntuación.

La comparación aportó además una enseñanza para la interpretación estadística: no detectar una diferencia no demuestra equivalencia. Las versiones de 30 y 40 preguntas tuvieron medias próximas, pero la investigación no estableció un margen de diferencia irrelevante ni desarrolló una prueba específica para confirmarlo. Del mismo modo, una diferencia significativa entre grupos requiere considerar su magnitud y el diseño, además del valor de p. Estas precisiones fortalecen la utilidad de los hallazgos para decisiones de investigación.

5.2.3. Limitaciones de la investigación

La primera limitación corresponde al muestreo no probabilístico. El reclutamiento por conveniencia y bola de nieve pudo concentrar determinados perfiles y redes de contacto. Los resultados describen a los participantes y no representan estadísticamente a todos los trabajadores ecuatorianos. La aplicación de cada versión en un período diferente añade una posible influencia del momento de recolección, que no puede separarse completamente de la condición de longitud.

La segunda limitación se relaciona con el instrumento. Las cuatro versiones procedieron de una adaptación del MLQ y conservaron contenido sobre liderazgo. No se evaluaron cuestionarios sobre otras variables administrativas, formatos de respuesta diferentes ni versiones inferiores a 30 preguntas. Asimismo, la reducción de reactivos modificó el conjunto de contenidos y controles, por lo que la comparación no equivale a añadir preguntas sin alterar ninguna otra característica de medición.

La tercera limitación afecta la construcción del índice. La regla de al menos una contradicción puede depender del número de pares; las categorías extremas y centrales no son siempre respuestas de baja calidad; y las formulaciones de atención de sentido opuesto requieren una orientación común de puntuación. Además, los porcentajes agregados del grupo de 30 preguntas no reproducen exactamente su media de calidad mediante una resta directa. La tabla de resultados conserva los valores del análisis, cuya reproducción requiere revisar los datos y la sintaxis de cómputo.

La cuarta limitación reside en la información temporal y el alcance de los registros. El tiempo total pudo incluir pausas, pero no se midieron directamente las interrupciones ni la multitarea. Los registros disponibles tampoco establecen el total de invitaciones, aperturas sin envío o no participantes. En consecuencia, el abandono registrado y las omisiones no deben convertirse en estimaciones completas de participación sin contar con esos denominadores.

Finalmente, la no normalidad, la desigualdad de tamaños y la falta de diagnósticos completos de los modelos aconsejan interpretar con cautela la inferencia. Las pruebas por pares se informaron sin un ajuste documentado por multiplicidad. Las regresiones excluyeron la constante y sus tablas no incluyeron intervalos de confianza ni medidas de ajuste. La precisión limitada de los coeficientes cuadráticos impidió establecer tiempos óptimos confiables. Estas condiciones delimitan los resultados y orientan las verificaciones para su ampliación.

CONCLUSIONES

La investigación permitió establecer que la extensión del cuestionario constituye un factor relevante en la calidad de respuesta de las encuestas digitales aplicadas en estudios administrativos. En los 473 trabajadores participantes, las versiones de 30 y 40 preguntas alcanzaron mayores medias de calidad que las versiones de 50 y 65 preguntas, encontrándose diferencias estadísticamente significativas entre los cuestionarios breves y los más extensos. Sin embargo, la ausencia de diferencias significativas entre las versiones de 30 y 40 preguntas, así como entre las de 50 y 65, evidencia que la relación entre longitud y calidad no sigue una disminución estrictamente proporcional. Por tanto, los resultados no permiten establecer un número universal de preguntas óptimo, sino que respaldan la necesidad de equilibrar extensión, cobertura conceptual y esfuerzo requerido al participante.

El análisis de los diferentes componentes de la calidad mostró que el incremento de la extensión estuvo acompañado principalmente por un aumento de las señales de desatención observada y de las respuestas lineales, mientras que indicadores como las omisiones, el abandono, los valores atípicos y el empleo de categorías extremas o centrales presentaron comportamientos menos uniformes. Asimismo, las versiones más extensas registraron coeficientes de consistencia interna superiores, pese a presentar menores medias en el índice global de calidad. Este resultado evidencia que la fiabilidad interna de un instrumento no debe utilizarse como único criterio para valorar la calidad de los datos, siendo necesario considerar conjuntamente indicadores de atención, consistencia, omisión, comportamiento de respuesta y condiciones de aplicación.

El tiempo de respuesta también se relacionó con la calidad de forma no lineal en las versiones de 30, 40 y 50 preguntas, lo que indica que una mayor duración no representa necesariamente una mejora continua de la calidad. De igual manera, la carga percibida y la edad mostraron asociaciones positivas con el índice de calidad en las cuatro versiones, mientras que el género femenino presentó asociación significativa en tres de ellas. Estas relaciones deben interpretarse como asociaciones estadísticas dentro de los grupos estudiados y no como evidencia de mecanismos causales o diferencias inherentes entre participantes.

Los resultados también ponen de manifiesto que la calidad de respuesta es un fenómeno multidimensional que no puede explicarse exclusivamente por la longitud del cuestionario. La variabilidad observada dentro de cada grupo, las diferencias en los indicadores específicos y las asociaciones encontradas con el tiempo, la carga percibida y determinadas características de los participantes sugieren la intervención simultánea de factores relacionados con el instrumento, el contexto de aplicación y la experiencia de respuesta. En consecuencia, futuras investigaciones deberían replicar la comparación mediante asignación aleatoria de las versiones, períodos equivalentes de aplicación y un núcleo común de reactivos y controles, además de incorporar cuestionarios con menos de 30 preguntas y diferentes contenidos administrativos. Esto permitiría determinar con mayor precisión qué variaciones de la calidad pueden atribuirse a la extensión y cuáles responden a las características particulares del instrumento o de la población estudiada.

A manera de cierre, el autor de la obra concluye que, el diseño eficiente de encuestas digitales requiere considerar la extensión como parte de un sistema más amplio de decisiones

metodológicas. La eficiencia no consiste únicamente en reducir el número de preguntas, sino en obtener información suficiente, interpretable y metodológicamente sólida con una demanda razonable para quien responde. En instrumentos administrativos semejantes a los evaluados, las versiones de aproximadamente 30 a 40 preguntas pueden constituir una referencia inicial favorable, siempre que preserven las dimensiones esenciales del fenómeno estudiado y que su funcionamiento sea corroborado mediante pilotaje y evaluación específica en cada población y contexto.

RECOMENDACIONES

Para aplicaciones semejantes, se recomienda examinar primero si una versión de 30 o 40 preguntas cubre adecuadamente los objetivos de medición. Cuando se necesiten instrumentos más extensos, conviene justificar el contenido adicional y evaluar su funcionamiento durante el pilotaje. La elección debe sostenerse en evidencia propia de la población y del cuestionario, sin adoptar el número de reactivos como único criterio de calidad ni eliminar preguntas esenciales solo para alcanzar una longitud predeterminada.

Se recomienda acompañar las puntuaciones sustantivas con duración, omisiones, patrones lineales y controles de atención o consistencia. Deben documentarse la dirección de los reactivos, los denominadores, el tratamiento de datos faltantes y las reglas del índice. En particular, la reproducción de esta investigación requiere verificar la recodificación de atención y la correspondencia entre los componentes y la puntuación final, manteniendo diferenciadas las señales observadas de las decisiones de exclusión.

Las nuevas investigaciones deberían comparar versiones durante períodos equivalentes y, cuando sea posible, asignar aleatoriamente a los participantes. También sería pertinente mantener un núcleo común de reactivos y controles para reducir diferencias de medición entre condiciones. La ampliación a cuestionarios de menos de 30 preguntas, otras variables administrativas y poblaciones distintas permitiría evaluar qué parte del patrón depende de la longitud y qué parte del instrumento o del contexto.

Para estudiar la duración óptima, se recomienda conservar coeficientes con precisión suficiente, examinar el rango de tiempos y presentar incertidumbre y diagnósticos de ajuste. La

medición de pausas o de tiempo activo permitiría distinguir deliberación de interrupciones. Asimismo, conviene contrastar modelos con constante y métodos adecuados a las distribuciones observadas, además de considerar la multiplicidad de las comparaciones y comunicar tamaños de efecto.

Finalmente, las asociaciones de carga, edad y género requieren replicación y explicaciones contrastables. Las futuras aplicaciones pueden diferenciar dificultad, cansancio y tedio, y medir experiencia con encuestas o familiaridad digital. Esas variables ayudarían a comprender los patrones sin atribuirlos anticipadamente a capacidades personales. La continuidad del estudio debe profundizar en la calidad del proceso de respuesta y conservar el objeto metodológico que dio origen a la investigación.

REFERENCIAS BIBLIOGRÁFICAS

American Association for Public Opinion Research (2023). Standard definitions: Final dispositions of case codes and outcome rates for surveys (10.ª ed.). https://aapor.org/standards-and-ethics/standard-definitions/

Andreadis, I., & Kartsounidou, E. (2020). The impact of splitting a long online questionnaire on data quality. Survey Research Methods, 14(1), 31–42. https://doi.org/10.18148/srm/2020.v14i1.7294

Artino, A. R., Jr., La Rochelle, J. S., DeZee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE guide no. 87. Medical Teacher, 36(6), 463–474. https://doi.org/10.3109/0142159x.2014.889814

Avolio, B. J., & Bass, B. M. (2004). Multifactor Leadership Questionnaire: Manual and sampler set (3.ª ed.). Mind Garden.

Bhaktha, N., Silber, H., & Lechner, C. M. (2026). A unified framework for characterizing response quality in surveys with multi-item scales. Communications Psychology, 4, Artículo 86. https://doi.org/10.1038/s44271-026-00463-2

Cernat, A., Sakshaug, J., Christmann, P., & Gummer, T. (2024). The impact of survey mode design and questionnaire length on measurement quality. Sociological Methods & Research, 53(4), 1873–1904. https://doi.org/10.1177/00491241221140139

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. Psychological Assessment, 31(12), 1412–1427. https://doi.org/10.1037/pas0000626

Cornesse, C., & Blom, A. G. (2023). Response quality in nonprobability and probability-based online panels. Sociological Methods & Research, 52(2), 879–908. https://doi.org/10.1177/0049124120914940

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5.ª ed.). SAGE.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. Journal of Experimental Social Psychology, 66, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Décieux, J. P. (2024). Sequential on-device multitasking within online surveys: A data quality and response behavior perspective. Sociological Methods & Research, 53(3), 1384–1411. https://doi.org/10.1177/00491241221082593

Décieux, J. P., & Sischka, P. E. (2024). Comparing data quality and response behavior between smartphone, tablet, and computer devices in responsive design online surveys. SAGE Open, 14(2), Artículo 21582440241252116. https://doi.org/10.1177/21582440241252116

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on

perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment, 29(2), 136–151. https://doi.org/10.1177/1073191120957102

Escurra, L. M. (1988). Cuantificación de la validez de contenido por criterio de jueces. Revista de Psicología, 6(1–2), 103–111. https://doi.org/10.18800/psico.198801-02.008

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current Practice and Recommendations. Social Psychological and Personality Science, 8(4), 370–378. https://doi.org/10.1177/1948550617693063

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. Public Opinion Quarterly, 73(2), 349–360. https://doi.org/10.1093/poq/nfp031

Garbarski, D., Dykema, J., Yonker, J. A., Bae, R. E., & Rosenfeld, R. A. (2025). Improving the measurement of gender in surveys: Effects of categorical versus open-ended response formats on measurement and data quality among college students. Journal of Survey Statistics and Methodology, 13(1), 18–38. https://doi.org/10.1093/jssam/smae043

Gibson, A. M., & Bowling, N. A. (2020). The effects of questionnaire length and behavioral consequences on careless responding. European Journal of Psychological Assessment, 36(2), 410–420. https://doi.org/10.1027/1015-5759/a000526

Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. The Leadership Quarterly, 31(4), Artículo 101384. https://doi.org/10.1016/j.leaqua.2020.101384

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. Public Opinion Quarterly, 74(5), 849–879. https://doi.org/10.1093/poq/nfq065

Gummer, T., Bach, R., Daikeler, J., & Eckman, S. (2021). The relationship between response probabilities and data quality in grid questions. Survey Research Methods, 15(1), 65–77. https://doi.org/10.18148/srm/2021.v15i1.7727

Izaguirre Olmedo, J. A., Jordán Correa, D. P., & Palacios Sarmiento, T. Y. (2025). Eficiencia en el tiempo de respuesta y calidad de los datos en encuestas en línea. Revista Venezolana de Gerencia, 30(Especial 13), 407–419. https://doi.org/10.52080/rvgluz.30.especial13.27

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology, 5(3), 213–236. https://doi.org/10.1002/acp.2350050305

Kunz, T., & Gummer, T. (2025). Effects of objective and perceived burden on response quality in web surveys. International Journal of Social Research Methodology, 28(4), 385–395. https://doi.org/10.1080/13645579.2024.2393795

Lawlor, J., Thomas, C., Guhin, A. T., Kenyon, K., Lerner, M. D., UCAS Consortium, & Drahota, A. (2021). Suspicious and fraudulent online survey participation: Introducing the REAL framework. Methodological Innovations, 14(3), Artículo 20597991211050467. https://doi.org/10.1177/20597991211050467

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. Psychological Methods, 23(3), 412–433. https://doi.org/10.1037/met0000144

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. Psychological Methods, 17(3), 437–455. https://doi.org/10.1037/a0028085

Nur, A. A., Leibbrand, C., Curran, S. R., Votruba-Drzal, E., & Gibson-Davis, C. (2024). Managing and minimizing online survey questionnaire fraud: Lessons from the triple c project. International Journal of Social Research Methodology, 27(5), 613–619. https://doi.org/10.1080/13645579.2023.2229651

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Schneider, S., Lee, P.-J., Hernandez, R., Junghaenel, D. U., Stone, A. A., Meijer, E., Jin, H., Kapteyn, A., Orriens, B., & Zelinski, E. M. (2024). Cognitive functioning and the quality of survey responses: An individual participant data meta-analysis of 10 epidemiological studies of aging.

The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 79(5), Artículo gbae030. https://doi.org/10.1093/geronb/gbae030

Sischka, P. E., Décieux, J. P., Mergener, A., Neufang, K. M., & Schmidt, A. F. (2022). The impact of forced answering and reactance on answering behavior in online surveys. Social Science Computer Review, 40(2), 405–425. https://doi.org/10.1177/0894439320907067

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. Psychological Bulletin, 133(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Vehovar, V., Couper, M. P., & Čehovin, G. (2023). Alternative layouts for grid questions in PC and mobile web surveys: An experimental evaluation using response quality indicators and survey estimates. Social Science Computer Review, 41(6), 2122–2144. https://doi.org/10.1177/08944393221132644

Wind, S. A., & Lugu, B. (2024). Combining nonparametric and parametric item response theory to explore data quality: Illustrations and a simulation study. Applied Measurement in Education, 37(2), 109–131. https://doi.org/10.1080/08957347.2024.2345592

World Wide Web Consortium (2024). Web Content Accessibility Guidelines (WCAG) 2.2. https://www.w3.org/TR/2024/REC-WCAG22-20241212/

Yu, E. C., Kopp, B., & Narine, V. R. (2026). Beyond survey length: Understanding respondent perceptions of

burden. Journal of Official Statistics, 42(3), 527–551. https://doi.org/10.1177/0282423x261451321

BIBLIOGRAFÍA 2026

PERFIL PROFESIONAL

Jorge Andrés Izaguirre Olmedo
Docente de la Universidad Internacional del Ecuador, (Ecuador).
joizaguirreol@uide.edu.ec
https://orcid.org/0000-0003-3231-8948

Economista con mención en Gestión Empresarial graduado de la Escuela Superior Politécnica del Litoral, (Ecuador). Magíster en Finanzas y Proyectos Corporativos graduado de la Universidad de Guayaquil, (Ecuador). DBA graduado de San Ignacio University, (Estados Unidos).

La obra sitúa en el centro de la reflexión metodológica una pregunta fundamental: ¿cómo recopilar información pertinente sin imponer una carga excesiva a quienes responden? La obra plantea un acercamiento a las relaciones entre la longitud de los cuestionarios, el tiempo requerido para completarlos y la calidad de la información obtenida, dentro de los desafíos que presenta la investigación administrativa en entornos digitales. Su enfoque invita a considerar cada encuesta como un instrumento cuya utilidad depende de decisiones coherentes con los objetivos del estudio y las características de los participantes. Desde esta perspectiva, la eficiencia comprende el equilibrio entre profundidad, claridad y viabilidad de la recolección de datos. El tema adquiere relevancia para investigaciones que buscan comprender percepciones, experiencias y comportamientos en organizaciones, donde disponer de información adecuada resulta esencial para sustentar análisis y decisiones. Dirigido a estudiantes, docentes, investigadores y profesionales de la administración, el libro propone una reflexión sobre el diseño de instrumentos digitales y su papel en la producción de conocimiento. Una contribución para pensar encuestas pertinentes, accesibles y orientadas a obtener datos útiles para el análisis administrativo.

ARTI EDITORIAL

ISBN: 978-9907-9600-1-3