



Dirección General de Servicio Civil

ÁREA DE RECLUTAMIENTO Y SELECCIÓN DE PERSONAL

**INFORME TÉCNICO
USEEI-INF-09-2022**

**Manual Orientativo en el diseño y construcción de
pruebas estandarizadas para evaluar competencias**

San José, Costa Rica
Diciembre, 2022

**Informe Técnico
 USEEI-INF-09-2022**

Nombre del Informe: Manual Orientativo en el diseño y construcción de pruebas estandarizadas para evaluar competencias		
Elaborado por: Jose Laurian Ramírez Díaz	Cargo: Profesional Analista	Fecha de elaboración: Octubre 2021
Revisado por: María Antonieta Chavarría Fernández Viviana Zamora Beita	Cargo: Profesional Analista Profesional Analista	Fecha de revisión: Octubre 2021 Octubre 2021 y Diciembre 2022
Aprobado por: Karla Carranza López David Campos Calderón	Cargo: Jefatura de la Unidad de Sistematización y Estandarización para la Evaluación de la Idoneidad Director del Área de Reclutamiento y Selección de Personal	Fecha de aprobación: Diciembre 2022 Diciembre 2022

Tabla de contenido

Introducción	5
Pruebas estandarizadas para evaluar competencias	6
1. Definición de los aspectos a medir	7
2. Validar la definición de los constructos o dimensiones y subdimensiones	10
3. Diseño y creación de ítems	15
4. Validación de ítems preliminares	17
5. Mejoramiento de ítems preliminares	19
6. Prueba Piloto	20
7. Validación estadística de ítems	21
- Confiabilidad por mitades	21
- Análisis factorial	23
- Índice de acuerdo interjueces	24
- Consideraciones generales para la recolección y el análisis estadístico de datos	25
8. Confección de prueba final	26
Bibliografía	29

Índice de tablas

Tabla 1. Ejemplo de constructos referidos a la inteligencia espacial en procesos de aprendizaje.....	8
Tabla 2. Ejemplo de dimensiones referidas a la variable Estrato Social.....	9
Tabla 3. Ejemplo de triangulación para validar constructos o dimensiones y subdimensiones en MS Excel.....	12
Tabla 4. Ejemplo de triangulación para validar ítems preliminares.....	19
Tabla 5. Escala de aceptación del índice de fiabilidad.....	22
Tabla 6. Criterios de rotación para el análisis factorial.....	24
Tabla 7. Apartados del Manual de Aplicación de la Prueba.....	27
Tabla 8. Apartados del informe técnico sobre el diseño de pruebas de evaluación de competencias.....	28

Índice de ilustraciones

Ilustración 1. Ejemplo de validación de un constructo.....	10
Ilustración 2. Ejemplo de validación de una dimensión o subdimensión.....	11
Ilustración 3. Distribución de las observaciones con base en un ACP.....	13
Ilustración 4. Análisis de correlaciones entre variables con base en el ACP.....	14
Ilustración 5. Gráfico lineal del coeficiente de correlación de Pearson.....	15

Introducción

El presente Informe Técnico tiene como objetivo brindar a las Oficinas de Gestión Institucional de Recursos Humanos (OGEREH) del Régimen de Servicio Civil, un paso a paso de las Etapas Metodológicas para diseñar y construir pruebas estandarizadas de evaluación de competencias.

Dicho Informe obedece en principio a la normativa establecida por la Dirección General de Servicio Civil que regula la realización de los Concursos Internos de los puestos cubiertos por el Régimen de Servicio Civil, respecto al establecimiento de las pruebas de competencias específicas como predictor para determinar la idoneidad.

Este Informe se sub divide en los siguientes ocho apartados, los cuales señalan de manera detallada cómo se diseña y construye una prueba estandarizada para la evaluación de competencias específicas, incluyendo definiciones y ejemplos, así como consideraciones de interés: Definición de los aspectos a medir, Validar la definición de los constructos o dimensiones y subdimensiones, Diseño y creación de ítems, Validación de ítems preliminares, Mejoramiento de los ítems preliminares, Prueba piloto, Validación estadística de los ítems y finalmente, Ensamblaje y versión final de la prueba.

Es importante aclarar que la información que aquí se estará aportando constituye una base para el diseño y creación de este tipo de pruebas. Sin embargo, cada OGEREH, tendrá que asumir la responsabilidad de considerar sus necesidades y condiciones propias, para la conformación de un equipo de profesionales que pueda realizar este trabajo.

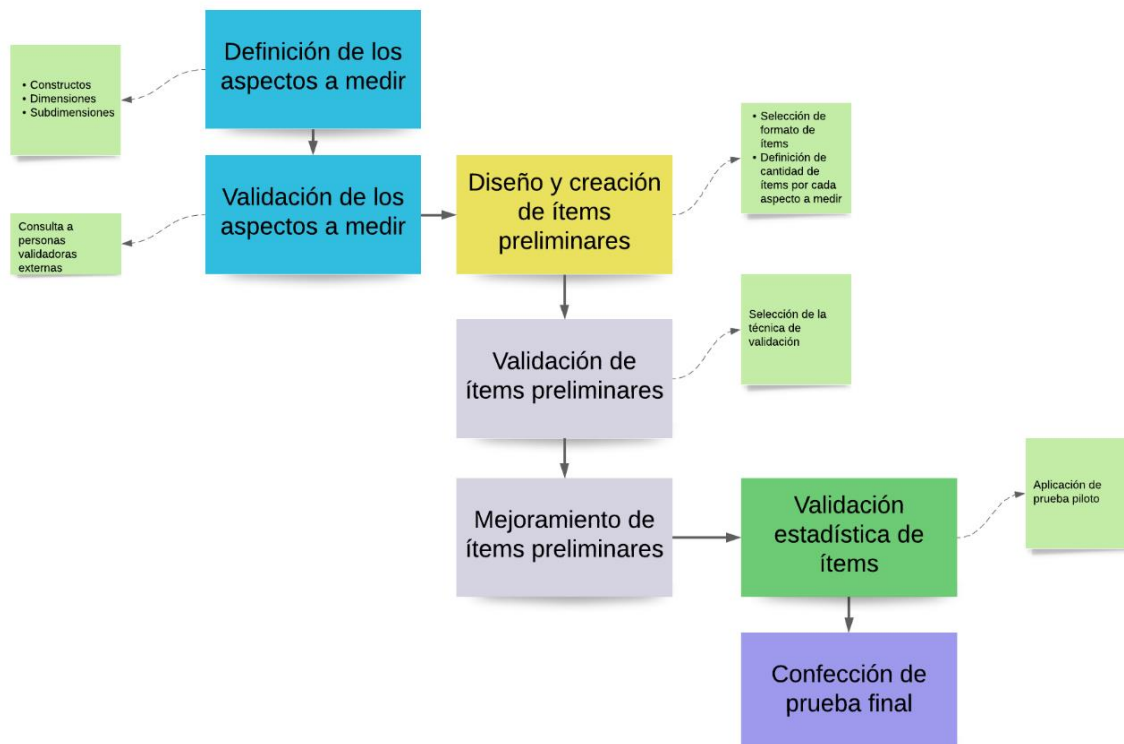
Pruebas estandarizadas para evaluar competencias

Una prueba consiste en una representación de ítems relacionados a lo que se desea medir, ya sea uno o más fenómenos, objetos, situaciones o personas y, para lo cual debe seguirse un procedimiento determinado. En otras palabras, una prueba mide o evalúa aspectos específicos con base en escalas de calificación e interpretación establecidas de previo, que están apegadas a un conjunto específico de normas o criterios (Medina, 2007).

Las pruebas estandarizadas para la evaluación de competencias tienen como objetivo la medición de aspectos comunes entre varias personas para obtener puntajes basados en escalas que permiten ubicarlas ordenadamente, según el resultado o desempeño que cada persona obtuvo.

Aunque existen varias formas para diseñar y construir estas pruebas, la metodología que aquí se presenta demostró su validez y confiabilidad al ser la utilizada por personas funcionarias del Área de Reclutamiento y Selección de Personal de la Dirección General de Servicio Civil para construir las pruebas de Evaluación de Competencias Transversales. De este modo y como punto de inicio, se muestra un diagrama con los pasos más importantes de la misma:

Figura 1. Aspectos metodológicos para diseñar y construir pruebas estandarizadas de evaluación de competencias



Nota: Elaboración propia (2021).

En las siguientes páginas se presentarán los aspectos metodológicos más relevantes para el diseño y creación de pruebas estandarizadas de evaluación de competencias con base en las etapas descritas en la figura anterior.

1. Definición de los aspectos a medir

La primera cuestión que debe resolverse al diseñar y construir una prueba estandarizada de competencias trata sobre definir cuáles variables van a medirse.

Es importante mencionar que una variable puede definirse como simple (un único aspecto a medir) o compleja (dos o más aspectos), y está referida a dos elementos básicos: 1) los criterios, cualidades, propiedades o características que se quieren evaluar y 2) los datos obtenidos para lograrlo (Cuestas, 2009; Villasís y Miranda, 2016; Soto, 2018). La carencia de una definición adecuada de variables y de datos robustos complicará el diseño y construcción de pruebas de evaluación para competencias específicas.

De esta forma, la definición de variables puede darse de distintas maneras. A continuación, se presentan dos de las formas más utilizadas para hacerlo:

- **Constructos:** son definiciones teóricas de las variables con base en comportamientos que no pueden definirse uno por uno, sino que deben relacionarse entre sí para medirlos adecuadamente (Soler, 2013). Un ejemplo son los test de personalidad que utilizan diversos criterios para medir los rasgos que distinguen a una persona de otra, como la introversión y la extroversión.
- **Dimensiones:** elementos disgregados de las variables complejas para su medición que, a diferencia de los constructos, se apegan a definiciones basadas en diversas fuentes (Soto, 2018). Dependiendo de su complejidad, podrían dividirse en unidades más pequeñas o subdimensiones. Un ejemplo son los exámenes clínicos de rutina, que en sí mismos miden aspectos en concreto pero que al integrar sus resultados determinan el estado global de salud de una persona.

Para ambos casos se recomienda construir una **definición literal** (a partir de fuentes bibliográficas), una definición operacional (cómo se entiende dicha variable en el contexto en el que se evaluará) y una definición comportamental (conductas a evaluar), tal como se verá más adelante.

A continuación, se presentan ejemplos de cómo definir las variables en constructos y dimensiones.

Ejemplos de la definición de Constructos

Ejemplo 1: Memoria espacial en procesos de aprendizaje

Tabla 1. Ejemplo de constructos referidos a la inteligencia espacial en procesos de aprendizaje.

CONSTRUCTO	DEFINICIÓN LITERAL	DEFINICIÓN OPERACIONAL	DEFINICIÓN COMPORTAMENTAL
Memoria espacial	Adquisición, retención y recuperación de conocimiento acerca de las características del entorno (Luna, Manzanares, Rodríguez y López, 2018, p.71).	Habilidad para recordar la ubicación de personas, lugares u objetos en lugares determinados.	<ul style="list-style-type: none"> - Recordar la distribución de aposentos ubicados en lugares determinados. - Recordar el orden de objetos según la ubicación. - Capacidad para reconocer lugares determinados a partir de elementos del paisaje. - Capacidad para reconocer personas con base en lugares determinados.
Espacio o entorno físico	Elemento mediador en la educación, en el aprendizaje y la socialización, que tiene connotaciones afectivas que influyen en las emociones (Huertas, 1988, citado por (González y Carmen, 1997).	Espacio físico en el que las personas interactúan entre ellas.	<ul style="list-style-type: none"> - Tipo de relaciones interpersonales en lugares determinados. - Normas sociales en lugares determinados. - Influencia de los valores personales en la interacción con otras personas. - Formas de interacción según el lugar en el que se encuentra la persona.

Nota: Elaboración propia (2021).

Como puede observarse, pese a que la inteligencia espacial en procesos de aprendizaje es la variable por medir y se podría hacer a partir de las dos capacidades indicadas en la definición comportamental (capacidad para recordar y reconocer), se ha agregado otro constructo relacionado al espacio o entorno físico. Nótese que, aunque podrían agregarse otros constructos, cada uno de los ejemplos planteados (Memoria espacial y Espacio o entorno físico) pueden medirse por separado.

Asimismo, nótese que la definición literal se basa en autores específicos, mientras que la definición operacional es una deconstrucción del constructo para adaptarlo al contexto específico de la evaluación.

Finalmente, la definición comportamental plantea las conductas más destacadas para el análisis y, tal como se verá más adelante, propone la relación directa entre el constructo y los ítems con los cuales será medido.

Es importante aclarar que si bien puede surgir la sensación de que algunos elementos quedaron por fuera, la inclusión o exclusión de estos se debe a una decisión que se toma en el camino, en aras de alcanzar la mayor unidimensionalidad¹ y homogeneidad de

¹ La unidimensionalidad se refiere a que un conjunto o grupo de ítems se basa en un único rasgo o constructo, por lo que las respuestas giran alrededor del mismo (Burga, 2006).

varianza² posibles, ya que esto influirá al seleccionar las técnicas de análisis estadístico, tal como se explicará más adelante.

A continuación, se expondrá un ejemplo para construir dimensiones.

Ejemplos de la definición de Dimensiones

Ejemplo 2: Dimensiones y subdimensiones de la variable Estrato Social

Tabla 2. Ejemplo de dimensiones referidas a la variable Estrato Social.

VARIABLE	DIMENSIONES	SUBDIMENSIONES
Estrato Social	Nivel de escolaridad de los adultos en el hogar	
	Nivel de escolaridad de los menores de edad en el hogar	
	Profesión de quienes trabajan en la familia	<ul style="list-style-type: none"> - Trabajo formal / informal - Trabajo propio / asalariado
	Condiciones de la vivienda	<ul style="list-style-type: none"> - Zona urbana / rural - Servicios públicos - Acceso a áreas de salud - Zonas verdes - Hacinamiento - Materiales de construcción

Nota: Adaptado de Bauce, Córdova y Avila (2018).

Como puede observarse, la construcción de dimensiones plantea un mayor grado de “libertad” al momento de definir los aspectos de la variable a medir. Nótese que los criterios en mención se alinean a tres cuestiones: por un lado, la variable Estrato Social se analiza desde una perspectiva etnográfica que considera la dimensión del nivel de escolaridad del sistema familiar de los adultos y los menores de edad en el hogar (primera y segunda dimensión respectivamente).

Por otro lado, la profesión de quienes trabajan se constituye en la tercera dimensión, la cual se desglosa en las subdimensiones de trabajo formal/informal y trabajo propio/asalariado y, finalmente, las Condiciones de la vivienda corresponde a la tercera dimensión y se divide en seis subdimensiones, las cuales están relacionadas a aspectos de tipo material y del contexto.

² La homogeneidad de varianza se refiere a la capacidad de una prueba para reproducir resultados parecidos entre poblaciones similares, es decir, que los resultados no variarán significativamente (Box, Hunter y Hunter, 2008).

2. Validar la definición de los constructos o dimensiones y subdimensiones

La acción de validar se relaciona con la eficacia o virtud de algo para lograr el efecto o resultado esperado. En este caso, la validez de los aspectos a medir está referida al grado de certeza que se tiene para evaluar lo que se desea con base en los constructos o dimensiones y subdimensiones que se han definido.

Para validar los constructos o dimensiones y subdimensiones es recomendable consultar a un número impar³ de personas que tengan conocimiento sobre los temas a tratar. Las mismas deberán evaluar la pertinencia de los aspectos considerados para diseñar y construir la prueba. Esta validación es preferiblemente cualitativa y se consideran criterios tales como:

- Claridad de la definición literal.
- Claridad de la definición operativa.
- Importancia del constructo o dimensión / subdimensión para evaluar la competencia.

Dicha evaluación puede basarse en una escala dicotómica (Adecuado / Inadecuado), tal como se muestra en los siguientes ejemplos:

Ilustración 1. Ejemplo de validación de un constructo.

Constructo	Definición literal	Definición operativa
Incertidumbre	La incertidumbre es la inseguridad que surge al contrastar la expectativas contra los posibles obstáculos que podrían evitar su logro (aspecto en el cual, la información de la que se dispone no siempre es cierta y precisa, pero que juega también un papel fundamental) (Restrepo P., 2017). De este modo, es un contrapuesto de la precisión y está referida a la imposibilidad de estimar el margen de error y cómo ello puede influir en la calidad de los productos finales (Ruiz, García y Mesa, 2010).	Diseñar estrategias o mecanismos para disminuir el margen de riesgo en los procesos, partiendo de una premisa de conocimiento.
	Adecuado	Inadecuado
Claridad de la definición literal	<input type="radio"/>	<input type="radio"/>
Claridad de la definición operativa	<input type="radio"/>	<input type="radio"/>
Importancia del constructo para l...	<input type="radio"/>	<input type="radio"/>

Nota: Elaboración propia (2021).

³ A modo de garantizar que, ante opiniones divididas, haya una mayoría (la mitad más uno) a favor de una postura que permita tomar decisiones.

Ilustración 2. Ejemplo de validación de una dimensión o subdimensión.

1.3. Disposición para la evaluación *

Subdimensión	Definición literal	Definición operativa
Disposición para la evaluación	La evaluación está relacionada a un proceso reflexivo y crítico de beneficio colectivo, que constituye una fuente de crecimiento e impulso para conocer en qué medida, se están cumpliendo los objetivos institucionales hacia el servicio ciudadano (evaluación), a lo interno de la institución o hacia sí mismas (autoevaluación) (Cruz y Quiñones, 2012).	Disposición hacia la reflexión crítica que permite valorar el cumplimiento de los objetivos planteados, tanto a lo interno de la institución como de las personas, así como la capacidad de autoevaluarse.

	Adecuado	Inadecuado
Claridad de la definición literal del constructo Asesorar y orientar	<input type="radio"/>	<input type="radio"/>
Claridad de la definición operativa del constructo Asesorar y orientar	<input type="radio"/>	<input type="radio"/>
Importancia del constructo del constructo Asesorar y orientar para la evaluación de la competencia	<input type="radio"/>	<input type="radio"/>

Nota: Elaboración propia (2021).

Es necesario aclarar que esta escala de validación puede tener más valores como pueden ser 1) Totalmente adecuado, 2) Adecuado e 3) Inadecuado. Incluso puede basarse en criterios de otra naturaleza como acuerdos, frecuencias, intensidades u otros, según el caso.

Por otro lado, para esta validación también se recomienda utilizar métodos o índices de medición que permitan calcular el nivel de concordancia entre los planteamientos teóricos y la definición de los constructos o dimensiones y subdimensiones. A continuación, se explican algunas alternativas para ello:

- **Acuerdo simple:** la forma más sencilla de hacerlo es sumar los valores y otorgarles un puntaje con base en la escala y el número de personas validadoras. Esto se puede hacer manualmente o por medio de hojas de cálculo, tal como se muestra en el ejemplo a continuación:

Tabla 3. Ejemplo de triangulación para validar constructos o dimensiones y subdimensiones en MS Excel.

A	B	C	D	E	G
	Preguntas	Observador 1	Observador 2	Observador 3	
Honestidad	1. Claridad de la definición literal	2	2	1	Adecuado
	1. Claridad de la definición operativa	2	2	2	Totalmente adecuado
	1. Importancia del constructo para la evaluación de la competencia	2	2	2	Totalmente adecuado
Liderazgo	2 Claridad de la definición literal	2	2	1	Adecuado
	2 Claridad de la definición operativa	2	2	2	Totalmente adecuado
	2 Importancia del constructo para la evaluación de la competencia	2	2	2	Totalmente adecuado

Nota: Elaboración propia (2021).

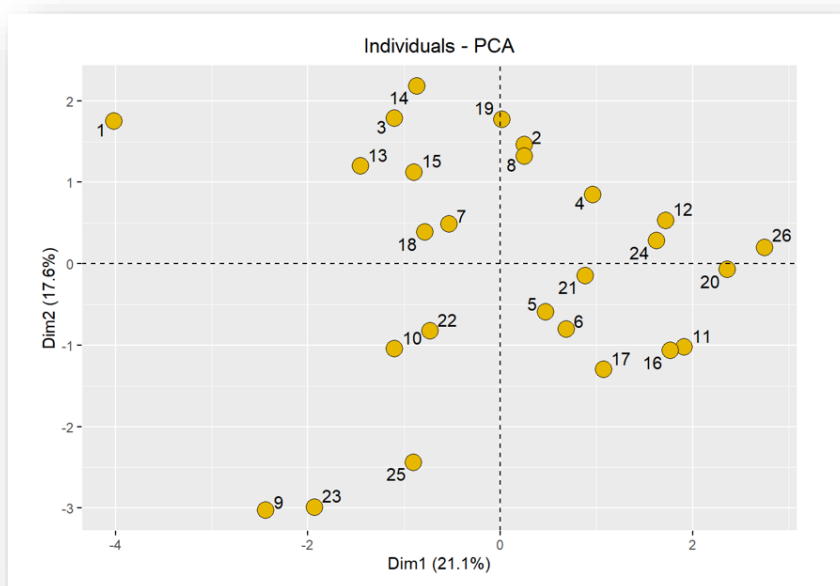
Para el ejemplo anterior se utilizó una hoja de cálculo en Microsoft Excel y se trabajó con una escala de tres valores: Totalmente adecuado (al cual se asignó un valor de 2), Adecuado (valor de 1) e Inadecuado (valor de 0). Al final se sumaron los puntajes (columna G) y se utilizó la siguiente codificación de tres colores:

- Verde (Totalmente adecuado): cuando todas las personas validadoras otorgaron el puntaje máximo, lográndose un total de 6 en las tres evaluaciones.
 - Amarillo (Adecuado): cuando el puntaje estuvo entre 3 a 5.
 - Naranja (Inadecuado): cuando el puntaje fue menor a 3.
- **Índice de acuerdos interjueces:** existen algunos índices para medir la fiabilidad de los acuerdos entre jueces (personas validadoras), entre los cuales destacan el Kappa de Cohen (dos jueces) o el Alpha de Krippendorff (tres o más jueces). Tanto

el Kappa como el Alpha en mención se distribuyen en una escala de 0 a 1, siendo los índices más cercanos a la unidad los que indican mayor nivel de acuerdo.

- **Análisis por Componentes Principales (ACP):** este método examina la relación y comportamiento de las variables y las observaciones (personas) que conforman la base de datos. En las siguientes páginas se muestran cuatro gráficos de un ACP a partir de las respuestas que dieron 26 personas sobre situaciones específicas de su trabajo.

Ilustración 3. Distribución de las observaciones con base en un ACP.

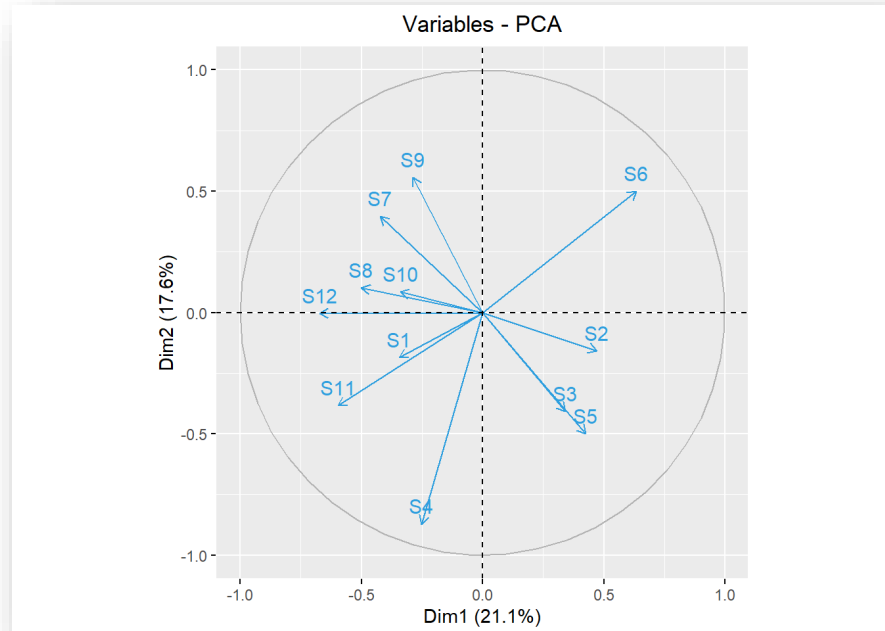


Nota: Elaboración propia (2021).

En el gráfico anterior puede verse la distribución de 26 observaciones o personas. Se destaca que, aunque no existió una homogeneidad óptima, las personas 1, 9, 23 y 25 mostraron un distanciamiento mayor al resto, lo que indica que sus respuestas fueron más heterogéneas.

En concordancia con lo anterior, la Ilustración 4 muestra la correlación de las respuestas dadas por las mismas personas analizadas en la Ilustración anterior.

Ilustración 4. Análisis de correlaciones entre variables con base en el ACP.



Nota: Elaboración propia (2021).

En este gráfico pueden considerarse las correlaciones fuertes negativas⁴ y las correlaciones fuertes positivas⁵.

Con respecto a las primeras, la variable S6 presenta una correlación fuerte negativa con la S1, S11 y S4. Esto quiere decir que el comportamiento de estas últimas tres variables es contrario al comportamiento de la S6, es decir, que en cuanto las primeras aumentan la otra disminuye, o viceversa. Asimismo, las variables S2, S3 y S5 mostraron el mismo comportamiento con relación a S9, S7, S8 y S10.

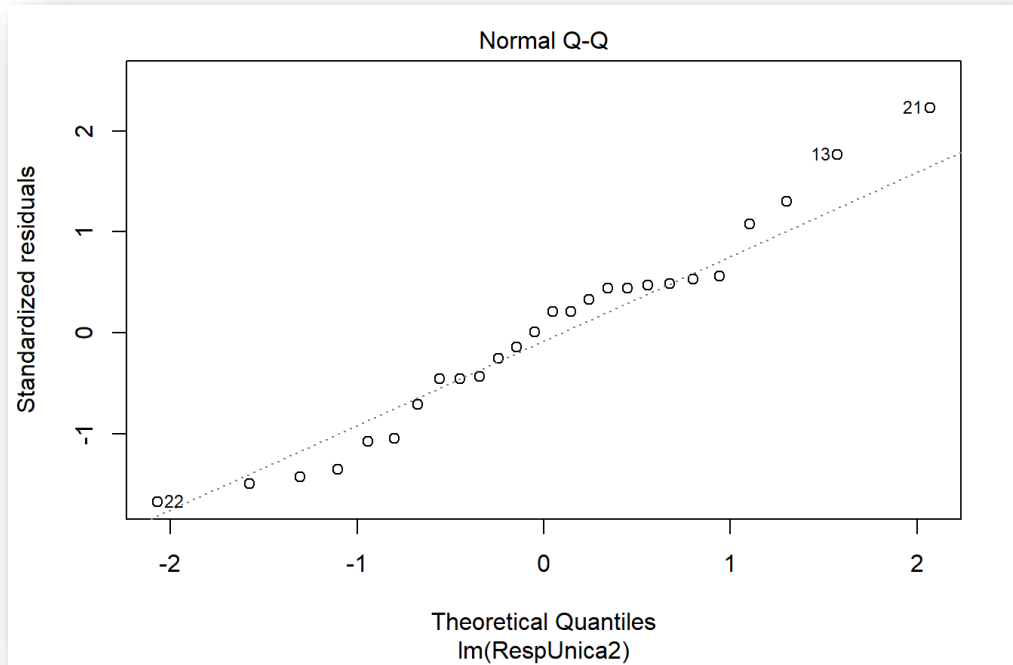
Al analizar las correlaciones fuertes positivas se encuentran tres agrupaciones principales: la primera de ellas entre las variables S1, S11 y S4. Esto quiere decir que el comportamiento de estas tres variables aumenta o disminuye de forma equivalente entre ellas. La segunda correlación fuerte positiva se puede ver entre las variables S2, S3 y S5 y, finalmente, la tercera agrupa a S9, S7, S8 y S10.

⁴ Las relaciones fuertes negativas indican que si una variable aumenta la otra disminuye, o viceversa. Ej. entre mayor desgaste presenta una llanta, su capacidad de fricción sobre la carretera será menor.

⁵ Las relaciones fuertes positivas indican que si una variable aumenta o disminuye la otra se comportará similar. Ej. entre más tiempo tarde una persona en la ducha mayor será el consumo de electricidad.

Finalmente, el gráfico lineal con base en el coeficiente de correlación de Pearson mostró lo siguiente:

Ilustración 5. Gráfico lineal del coeficiente de correlación de Pearson.



Nota: Elaboración propia (2021).

Puede observarse el comportamiento de las respuestas con base en la línea central que muestra la normalidad de los datos. En este caso, las observaciones 13, 21 y 22 están fuera de lo esperado, de modo que pueden interpretarse como datos atípicos o “cisnes negros”, y deben analizarse con mayor detalle para el tratamiento respectivo.

Es importante ofrecer dos acotaciones a esta altura del documento: primero, los ejemplos presentados fueron análisis llevados a cabo con “R Studio” (programa estadístico) con propósitos ilustrativos, de modo que presentan generalidades sin pretender ser exhaustivo y, segundo, se recomienda que el análisis e interpretación de datos sean realizados o acompañados por personas que conozcan sobre la temática, ya que las interpretaciones y resultados erróneos pueden tener serias repercusiones en este sentido.

3. Diseño y creación de ítems

Una vez que se han validado los constructos o dimensiones y subdimensiones, el siguiente paso consiste en diseñar y crear los ítems, los cuales constituyen la materia prima de las pruebas de evaluación de competencias. Para este paso es sumamente útil considerar las

conductas específicas integradas en la definición comportamental de los constructos o dimensiones y subdimensiones.

Con respecto al diseño de los ítems, es importante indicar que en esta etapa se toman decisiones sumamente relevantes tales como el formato de los ítems (Ej. Selección única, escalamiento de Likert, escalas dicotómicas, entre otros) y la cantidad de los mismos.

Con respecto a la cantidad de los ítems es importante indicar que, durante la validación estadística, la tasa de “pérdida” oscila entre el 40% al 60%. Esto quiere decir que por cada 10 ítems podrían suprimirse entre cuatro a seis (a veces más), de modo que su número debe ser suficiente para afrontar este paso.

Una vez que se han tomado las decisiones relacionadas al diseño y formato de los ítems, el siguiente paso es construirlos. Si bien no existe una única forma para hacerlo y este paso demanda distintas habilidades de redacción, vocabulario, capacidad de síntesis entre otras, Morales, Urosa y Blanco (2003) plantean algunas pautas para dicho proceso:

- **Redacción.** Para evaluar competencias generalmente se utilizan ítems de dos tipos: los primeros tienden a plantear opiniones constituidas en actitudes verbalizadas, en las que las personas pueden estar de acuerdo o no, mientras que los segundos pueden redactarse en forma de conductas o rasgos personales.
- **Relevancia.** Los ítems deben relacionarse con la variable que se desea medir.
- **Claridad.** Al construir ítems una de las consideraciones que debe tomarse en cuenta es que las personas que vayan a contestar las pruebas comprendan su contenido de forma similar. Para esto se hacen dos recomendaciones generales: 1) evitar las expresiones negativas, especialmente la palabra “no”, y 2) que cada ítem incluya una sola idea, evitando el uso de conjunciones tales como “o” e “y”.
- **Discriminación.** Los ítems deben redactarse de forma que haya personas que estén de acuerdo y otras que no. Debe cuidarse de que el ítem no sea muy transparente (que la respuesta sea obvia para la mayoría), y que tampoco recaiga en la deseabilidad social (que propongan situaciones en donde las personas se esmeren por dar la mejor imagen de sí mismas en lugar de responder sinceramente).
- **Repetición.** La fiabilidad para medir una variable va de la mano con el número de ítems que preguntan sobre lo mismo. Por tanto, se deben formular ítems que pregunten sobre la misma idea con el cuidado de que las personas no los identifiquen fácilmente.

- **Bipolaridad.** En una prueba deben existir ítems en ambas direcciones, es decir, que unos planteen situaciones positivas y otras inversas o negativas, entendiéndose que para los segundos no se trata de utilizar palabras como “no”, sino que planteen situaciones opuestas respecto a otros (Ej. “Estudiar es una actividad divertida” / “Estudiar es aburrido”). Asimismo, es recomendable que exista un balance entre ambos tipos, preferiblemente por mitades (Ej. Cinco ítems positivos y cinco ítems inversos o negativos).

4. Validación de ítems preliminares

Cuando se han creado los ítems, los mismos deben someterse a un proceso de validación para asegurar de que las personas comprenden lo que se les está planteando en cada caso. Para ello, se recomienda utilizar técnicas tales como la entrevista cognitiva, la cual es posible de realizar tanto presencial como virtualmente, o estrategias como los grupos focales⁶.

La Entrevista Cognitiva consiste en una evaluación cualitativa que se utiliza en investigaciones psicológicas para validar ítems preliminares mediante la comprensión de fenómenos psicosociales (Willis, 2015). La misma se ha convertido en una estrategia que permite identificar problemas de comprensión y claridad entre las preguntas y respuestas al construir pruebas psicométricas (Miller, Willson, Chepp & Padilla, 2014).

La entrevista cognitiva consiste en mostrar los ítems preliminares a un grupo de personas con un perfil similar a la población que será evaluada con la prueba (preferiblemente a un mínimo ocho personas), para que valoren la comprensibilidad y la claridad del vocabulario, a modo de garantizar que su planteamiento sea adecuado para medir lo que se requiere. Para ello, existen algunas pautas generales para el desarrollo de esta, tales como:

- **Saludo y presentación.** Se ofrece un encuadre de la actividad en el que se indican aspectos tales como los objetivos, la metodología, el tiempo estimado para el desarrollo de la sesión, entre otros.
- **Ejercicio de Visualización Guiada.** Se recomienda hacer una actividad de imaginación guiada para relajación y concentración de previo al inicio de la actividad.
- **Entrega de cuestionario o presentación de ítems.** Consiste en mostrar los ítems y solicitar a la(s) persona(s) que valore(n) cada uno con base en los criterios que se explicarán más adelante.

⁶ “Un espacio de opinión para captar el sentir, pensar y vivir de los individuos, provocando auto explicaciones para obtener datos cualitativos” (Hamui y Varela, 2012, p.56).

- **Lectura en voz alta de cada ítem.** Es importante que la(s) persona(s) lea(n) el ítem en voz alta para anotar todos los aspectos que puedan entrecerse de los mismos (Ej. Palabras con mayor dificultad al momento de leerlas, pausas, gestos del rostro, entre otros). Cada vez que se perciba alguna reacción que sea importante de anotar, debe consultársele a la(s) persona(s) a qué se debe.
- **Sistematización o registro.** Consiste en la bitácora de la entrevista cognitiva, en la cual se registran los aspectos observados más relevantes para cada ítem, de modo que pueda consultarse la información cuantas veces sea necesario.
- **Grabación.** El uso de herramientas tecnológicas se ha convertido en un *per se* de las actividades cotidianas. Es recomendable grabar cada entrevista cognitiva, previo Consentimiento Informado⁷, a modo de contar con un registro para repasar aspectos relevantes cuando se requiera, indistintamente a que la misma haya sido presencial o virtual.

Con respecto a la valoración de los ítems, la misma debe enfocarse sobre aspectos que permitan identificar mejoras, tales como los descritos a continuación:

- Redacción y comprensión:
 - ¿El ítem es comprensible?
 - ¿La redacción del ítem permite una lectura fluida?
- Vocabulario y claridad:
 - ¿El vocabulario es claro?
 - ¿Las palabras utilizadas en el ítem son adecuadas?
- Sugerencias para mejorar el ítem:
 - ¿Cómo mejoraría el ítem para una comprensión mejor del mismo?
 - ¿Cuáles palabras sustituiría para mejorar su lectura?

Una vez que se hayan mostrado todos los ítems, es importante recabar la impresión de cada una de las personas que participaron en la Entrevista Cognitiva para realimentar el proceso. Para ello se podría preguntar:

- ¿Qué le pareció la experiencia en general?

⁷ Para el desarrollo de esta y otras estrategias o técnicas, es necesario contar con el Consentimiento Informado de quienes participan, en el cual da fe de que se ha informado adecuadamente sobre las condiciones referidas a la participación en el estudio, así como el uso de la información.

Para analizar la información de las entrevistas cognitivas, se sugiere el empleo de las técnicas descritas en este documento en lo referente a los constructos o dimensiones y subdimensiones, a saber:

- Acuerdo simple.
- Índice de acuerdos interjueces.
- Análisis por Componentes Principales.

A continuación, se muestra un ejemplo haciendo uso de la técnica de Acuerdo Simple, para el cual se recomienda crear una matriz en una hoja de cálculo con fórmulas programadas para obtener un puntaje para cada ítem con base en los aportes de todas las personas que los validaron:

Tabla 4. Ejemplo de triangulación para validar ítems preliminares.

Vocabulario y Claridad	Sugerencias	Redacción y comprensión	Sugerencias	Resumen de Observaciones	%
Debe desarrollarse a qué se refiere la palabra "filialmente".		Indica que la palabra "filiales" la relaciona con asociaciones futbolísticas.		La palabra filialmente debe cambiarse, y corregir palabras que son muy extensas.	66,67
Claro.	Es importante que se conceptualice qué se entiende por Probidad.	No comprende el ítem ya que el principio de probidad es lo más alto de lo que se refiere a ambas opciones de respuesta.	No le quedan claras las opciones de respuesta, ya que no reflejan el principio de probidad que él tiene.	La opción de respuesta no queda clara. Es conveniente indicar qué se entiende por probidad.	45,45
Claro.		Claro.			93,94

Nota: Elaboración propia (2021).

Para este caso, cuando algún ítem no haya recibido alguna recomendación por parte de las personas que validaron, obtendría una nota perfecta (Ej. 100 de 100). En caso de que haya observaciones, cada una le restaría puntos.

5. Mejoramiento de ítems preliminares

Considerando el ejemplo de la Tabla 6, el mejoramiento de los ítems se basa en la sistematización de la información recopilada en las Entrevistas Cognitivas (o la técnica que se haya seleccionado).

Cuando el aspecto validado (redacción, comprensión, claridad o sugerencias para mejorar el ítem) estuvo bien para la persona que, valida dichos ítems, es decir, que no recomienda mejoras, se suma una unidad (1) en el espacio correspondiente. Para los casos en los que se hagan sugerencias o comentarios se pone un cero (0), se marca la celda con color amarillo para visualizarlos fácilmente y se suma el total de cada ítem para obtener el porcentaje respectivo.

Una vez sistematizadas las observaciones y evaluados todos los ítems se cuenta con los insumos para mejorarlos. Es importante aclarar que este tipo de procesos conlleva un análisis constante, ya que podrían encontrarse sugerencias que no sean factibles de adaptar, de modo que el equipo de trabajo debe tomar decisiones en todo momento sobre la forma en cómo adaptará las observaciones o los comentarios de las personas validadoras.

Aunque existen diversas estrategias y métodos para ello, el mejoramiento de los ítems preliminares puede considerarse con base en ciertos criterios. A continuación, se plantean algunos puntos orientativos para este proceso:

- Varias personas muestran inquietud o disconformidad sobre un mismo aspecto.
- Ítems que plantean situaciones o conductas sin definir un contexto específico.
- Palabras con una interpretación distinta o contraria a la idea original del ítem.
- Redistribución de las ideas del ítem para una mejor comprensión o lectura.
- El ítem es muy extenso o presenta varias ideas que lo hacen confuso.
- Palabras cuya interpretación es muy diversa o desconocidas.
- Uso innecesario de un lenguaje excesivamente técnico.
- Falta de un lenguaje debidamente inclusivo o accesible.
- Lenguaje poco común o palabras rebuscadas.

Es recomendable que estas mejoras se hagan en equipo, es decir, que intervengan varias personas para tomar la decisión correspondiente.

6. Prueba Piloto

Una vez establecidos los ítems preliminares, con sus mejoras respectivas, se deben someter a una Prueba Piloto, en la cual se identifican las siguientes fases:

6.1 Preparación de la Prueba Piloto

La preparación de la Prueba Piloto se refiere a todos los aspectos de logística que permiten implementar la misma; a continuación, se identifican los siguientes pasos relevantes:

- Ensamblar la Prueba Piloto.
- Establecer el método de recolección de las respuestas para su calificación.

- Asignar al personal profesional a cargo de la aplicación de las mismas, para definir la organización de los grupos.
- Identificar la muestra (características y cantidad) a la que se le aplicará la Prueba Piloto.
- Gestionar los permisos correspondientes con otras instituciones para la autorización de las personas participantes.
- Establecer lugar y fechas de aplicación.
- Organizar los grupos para aplicación de la prueba, según fecha, lugar y su capacidad.
- Enviar convocatoria a las personas participantes, con un tiempo prudencial para considerar las reprogramaciones, así como posibles motivos de ausentismo.

6.2 Aplicación de la Prueba Piloto

La Prueba Piloto, consiste en la aplicación en sitio de esta a la muestra establecida, procurando contar con todas las condiciones y exigencias que se tendrían en la prueba real.

Con ella se pretende validar todos los aspectos asociados a esta, tales como: comprensión de los ítems, tiempo y métodos para su ejecución y calificación, entre otros, con los que se pueden realizar análisis estadísticos para tomar decisiones definitivas para la prueba final.

7. Validación estadística de ítems

Cuando se cuenta con los ítems mejorados (posterior a la revisión a causa de las Entrevistas Cognitivas), estos deben validarse estadísticamente, a modo de seleccionar aquellos que midan efectivamente lo que se desea. Para este propósito se recomiendan algunas técnicas tales como el análisis de confiabilidad por mitades, el análisis factorial o el cálculo de la fiabilidad del juzgamiento por personas expertas, las cuales pueden hacerse con programas estadísticos, los cuales se explican a continuación.

- **Confiabilidad por mitades**

Esta técnica evalúa la consistencia interna y requiere que las personas respondan una sola vez a la prueba. Se recomienda su uso para casos en los que se deban evaluar muchas variables y/o que las mismas sean heterogéneas, y conlleva los siguientes pasos:

- a) Se analizan los ítems según los constructos o dimensión a la que pertenecen. Para ello se analizan dos índices:
 - El aumento en el coeficiente del α (Alpha) de Cronbach⁸ cuando se suprime un ítem (para este cálculo se suprime uno por uno).

⁸ El coeficiente del α (Alpha) de Cronbach mide la homogeneidad o consistencia interna de una prueba calculada a partir de la correlación de cada ítem con los demás (Quero, 2010).

- El índice de correlación de los ítems, en donde se intenta encontrar las mejores medidas en este sentido, va de la mano con el cálculo del coeficiente del α (Alpha) de Cronbach.
- b) Cuando se han identificado los mejores ítems, ya sea porque el coeficiente del α (Alpha) de Cronbach no aumenta más o porque las correlaciones entre estos son las mejores posibles, se divide la prueba en dos mitades para calcular la fiabilidad. Para ello se recomienda hacer esta división con base en ítems pares para una mitad e impares para la otra.
- c) Una vez estimado el índice de fiabilidad de cada mitad con el α (Alpha) de Cronbach, se calcula la correlación entre ambas puntuaciones. Para ello se utiliza la siguiente fórmula:

$$r' = \frac{r_1 + r_2}{2}$$

Donde r_1 corresponde a la fiabilidad de la mitad de ítems pares y r_2 a la fiabilidad de los ítems impares.

- d) Una vez que se conoce la fiabilidad de cada mitad (r_1 y r_2) se calcula la fiabilidad de la prueba en su totalidad con base en la fórmula profética de Spearman-Brown (Morales, Urosa y Blanco, 2003; Barranza, 2007). Para ello se unen ambas mitades con los mejores ítems y se aplica la siguiente fórmula:

$$n = \frac{2r'}{1 + r'}$$

Donde r' es el índice de fiabilidad alcanzado al calcular el grado de correlación entre ambas mitades, considerando que las escalas de aceptación son:

Tabla 5. Escala de aceptación del índice de fiabilidad.

Índice	Interpretación
Menor a 0,7	Inaceptable
Mayor o igual que 0,7 pero menor que 0,8	Aceptable
Mayor o igual que 0,8 pero menor que 0,9	Adecuado
Mayor o igual 0,9	Óptimo

Nota: Elaboración propia (2021).

De esta forma, los considerandos anteriores plantean la metodología general para analizar la fiabilidad por mitades. A continuación, se presentan los aspectos que se deben tomar en cuenta para llevar a cabo un análisis factorial exploratorio.

- **Análisis factorial**

El análisis factorial, es definido por García, Gil y Rodríguez (2019) de la siguiente forma:

El análisis factorial tiene como propósito descubrir las diferentes dimensiones de variabilidad común existente en cierto campo de fenómenos que se hace operativo a partir de un grupo de variables (...) para reducir la información relativa a un conjunto inicial de variables describiendo las relaciones que las unen a partir de un modelo más simple (pp.12-13).

Existen dos tipos de análisis factorial: el exploratorio y el confirmatorio. En este documento se hace referencia al primero (Análisis Factorial Exploratorio, en adelante AFE), el cual tiene como objetivo la identificación del modelo más efectivo para medir el fenómeno, extrayendo los factores y, a diferencia del análisis por mitades, esta técnica de análisis utiliza cálculos estadísticos para agrupar los ítems a partir del comportamiento de las respuestas, de modo que los factores pueden variar con respecto a los constructos o a las dimensiones y subdimensiones iniciales, es decir, que la agrupación final de los ítems podría variar ya que algunos presentan una homogeneidad distinta a la agrupación que se asignó al inicio.

Esta técnica es recomendable para pruebas con variables homogéneas o de número reducido (Ej. tres o cuatro variables), y se emplea para identificar el modelo más efectivo de medición en donde, por ejemplo, si se plantean tres dimensiones de 20 ítems cada una (60 ítems en total), el AFE podría demostrar que con siete ítems en cada una de las tres dimensiones se logra una fiabilidad igual o superior que con los ítems iniciales. Es decir, con los resultados del AFE se pueden identificar los 21 ítems que mejor explican o miden las tres variables en lugar de 60.

Otro aspecto importante de considerar para el caso del AFE son los criterios de rotación, con los cuales “se busca una mayor simplicidad en los factores, en las variables o en ambas cosas a la vez” (García, Gil y Rodríguez, 2019, p.28) y que se basan en los tres criterios mostrados a continuación:

Tabla 6. Criterios de rotación para el análisis factorial.

Criterio de rotación	Alcance
Quartimax	Reduce los factores a partir de cada variable.
Varimax	Reduce las variables según los factores.
Equamax	Simplifica los factores y las variables simultáneamente.

Nota: García, Gil y Rodríguez (2019).

Cada factor debe someterse al análisis de fiabilidad respectivo, considerándose la escala de aceptación descrita en la Tabla 7 de modo que, en caso de encontrarse con un índice inaceptable, debe considerarse la supresión del mismo.

- Índice de acuerdo interjueces

Una de las características de las mediciones anteriores (análisis por mitades y AFE) es que se requiere de una muestra representativa para poder hacer cálculos adecuados, sobre lo cual se hablará más adelante.

La medición del índice de acuerdos interjueces es muy útil cuando se cuenta con una muestra reducida (menor a 50 personas, por ejemplo), y una de las técnicas más utilizadas para ello es el α (Alpha) de Krippendorff. Sin embargo, para utilizarla se recomienda analizar algunos aspectos previos:

- a) Se debe agotar la posibilidad de validar la prueba con el mínimo de personas requerido para cualesquiera de los métodos anteriores (análisis por mitades o AFE).
- b) Contar con personas expertas en las temáticas o variables evaluadas. Para ello, el mínimo recomendado son tres personas, aunque lo ideal es contar con 10 o más.
- c) Contar con una definición adecuada de los rubros que se evaluarán, tomando en cuenta que todos los ítems deben validarse por igual para identificar los mejores de cada constructo o dimensión correspondiente.
- d) El α (Alpha) de Krippendorff no evalúa la fiabilidad de los ítems, sino que calcula la fiabilidad con la que las personas validadoras o jueces coinciden en que cada uno de ellos logra medir la variable en cuestión y que se utilice o no en la prueba.

De esta forma, el α (Alpha) de Krippendorff es un índice entre 0 a 1 que mide la fiabilidad del acuerdo interjueces según la siguiente escala.

Escala de interpretación del α (Alpha) de Krippendorff.

Índice	Interpretación
Menor a 0,67	Inaceptable
Mayor o igual que 0,67 pero menor que 0,8	Adecuado
Mayor o igual que 0,8 pero menor que 0,99	Óptimo
Igual que 1	Total

Nota: Neuendorf (2017).

Es importante indicar que la Unidad de Sistematización y Estandarización para la Evaluación de la Idoneidad (USEEI), cuenta con una herramienta en Excel cuyo uso es muy sencillo, la cual se encuentra a disposición de las OGEREH y puede coadyuvar con el cálculo de este índice cuando así se requiera.

- **Consideraciones generales para la recolección y el análisis estadístico de datos**

Una de las consideraciones más importantes que deben tomarse en cuenta a la hora de recolectar datos de pruebas piloto está referida al establecimiento de pautas logísticas que favorezcan el resguardo de la confidencialidad de la información, ya que una fuga de datos podría perjudicar todo el proceso realizado.

Con respecto a la recolección de la información, este puede hacerse mediante la aplicación de cuestionarios físicos, así como por formularios electrónicos. Se recomienda el empleo de la segunda alternativa por varias razones: 1) por cuestiones de sostenibilidad ambiental al ahorrar recursos como la tinta y el papel, 2) porque el almacenamiento de la información es sincrónico e inmediato, 3) porque se maximiza el tiempo al registrar las respuestas tal como las personas han contestado y 4) se evitan los errores que pueden cometerse al digitar la información manualmente.

Es importante mencionar que existen varias alternativas en el mercado para crear formularios en línea, aunque el resguardo de la confidencialidad de la información está garantizado solamente con la contratación o pago de planes con las empresas que lo proveen, un aspecto que debe ser tomado en cuenta para la toma de decisiones correspondientes.

Otro punto que debe considerarse es que los métodos de análisis aquí mencionados (análisis por mitades, AFE e índice de acuerdo interjueces) pueden realizarse por medio de programas estadísticos, entre los cuales destacan el [Statistical Package for the Social Sciences \(SPSS\)](#) o [The R Project for Statistical Computing \(R\)](#), de licencia privativa y de código abierto respectivamente. Es importante indicar que el análisis de la información debe estar a cargo de personas debidamente capacitadas para ello.

Con respecto al tamaño de la muestra para aplicar la prueba piloto, definida como “un subconjunto de unidades estadísticas extraído del universo del cual se quiere conocer ciertas características” (Clairin y Brion, 2001, p.14), existen varias formas de calcular el número de personas necesario. García, Gil y Rodríguez (2019) plantean dos posturas teóricas para ello:

- La primera de ellas se basa en el número de ítems que deben validarse estadísticamente, donde el número de personas debe doblar o triplicar al número total de ítems o, idealmente, contar con cinco personas por cada uno. Existen otras posturas que hablan de una relación de 10 participantes por cada ítem.
- La segunda establece un mínimo de personas por prueba, estimando muestras de entre 400 a 150 personas como mínimo, aunque algunos autores consideran que el mínimo muestral debe ser 100. Es importante recordar que entre mayor sea el número de la muestra, menor será el error típico de la prueba.

Finalmente, es importante prestar atención al aumento o disminución del índice de fiabilidad de cada constructo o dimensión y subdimensión. La literatura indica que para los factores cuyo Alpha sea menor a 0,7 deberán suprimirse, ya que su capacidad de medición no es adecuada.

8. Confección de prueba final

Una vez que se han analizado los ítems con base en los datos obtenidos en la prueba piloto, se pueden identificar los mejores ítems, de esta manera se realiza el último paso para la confección de la prueba final.

Es importante considerar que, además de los ítems y los factores o grupos que conforman, es necesario crear los siguientes documentos:

- Manual de aplicación de la prueba, en el cual se indican aspectos relevantes tales como los que presentan a continuación:

Tabla 7. Apartados del Manual de Aplicación de la Prueba.

Apartado	Descripción
Caracterización de la prueba	Introducción en la que se explica las características más relevantes de la prueba.
Aplicación	Se indican aspectos tales como: <ul style="list-style-type: none"> - El tiempo y las formas de aplicación (física o electrónica). - La consigna de aplicación (partes de la prueba y ejemplo de los ítems).
Contenido específico de la prueba	Se indican los constructos o dimensiones y subdimensiones, o los factores que constituyen la prueba, según corresponda.
Calificación	Se indica cómo calcular el puntaje de la prueba. Es importante distinguir los ítems positivos de los negativos para evitar confusiones.
Materiales para la aplicación	Descripción de los requerimientos para aplicar la prueba.
Población a la que va dirigida la prueba	Caracterización del perfil o requisitos de las personas a las que se aplica la prueba.
Perfil de la persona aplicadora de la prueba	Descripción de las características idóneas de quién debe aplicar la prueba.
Criterios de validez y confiabilidad de la prueba	Presenta la fundamentación científica que sustenta la prueba a nivel teórico y metodológico (cualitativo y estadístico).
Recomendaciones	Describe pautas o prácticas que deben acompañar la aplicación de la prueba.
Fuentes consultadas	Referencia a códigos, normativa, criterios, asesorías u otros que hayan sido considerados para el diseño y creación de la prueba.
Fuentes bibliográficas	Descripción de las fuentes que corresponden a las citas o referencias que se hayan utilizado en el documento, con base en el formato de estilo APA vigente.
Aplicativo de la prueba	Se anexa el aplicativo de la prueba (los ítems).

Nota: Elaboración propia (2021).

- Informe técnico sobre el diseño de pruebas de evaluación de competencias, en el cual se describe el proceso de investigación que se dio durante el diseño y construcción de la prueba y en el que se indican aspectos tales como:

Tabla 8. Apartados del informe técnico sobre el diseño de pruebas de evaluación de competencias.

Apartado	Contenido
Introducción	<ul style="list-style-type: none"> - Antecedentes y justificación. - Delimitación del problema. - Objetivo General. - Objetivos específicos.
Marco conceptual	Se presenta la teoría que sustentó la investigación y a los componentes principales de la(s) prueba(s).
Marco metodológico	<ul style="list-style-type: none"> - Tipo de Investigación. - Etapas de la investigación documental. - Validación de los constructos o dimensiones y subdimensiones. - Aspectos sobre la redacción de los ítems. - Aplicación de las entrevistas cognitivas o grupos focales. - Preparación de la Prueba Piloto. - Aplicación de la Prueba Piloto. - Análisis de los datos.
Conclusiones y recomendaciones	Síntesis de los principales hallazgos y buenas prácticas para replicar en procesos futuros.

Nota: Elaboración propia (2021).

Es sumamente importante que se creen este tipo de documentos con el fin de sistematizar los procesos, a modo de memoria institucional y como evidencia de la excelente labor que se lleva a cabo en la Administración Pública. Asimismo, son un respaldo necesario en caso de alguna apelación o reclamo de las personas usuarias, ya que se trata de documentar las bases y fundamentos de la prueba.

Bibliografía

- Barranza, A. (2007). ¿Confiabilidad? *Apuntes sobre metodología de la investigación*, (6), 6-10.
- Bauce, G., Córdova, M. y Avila, A. (2018). Operacionalización de variables. *Revista del Instituto Nacional de Higiene "Rafael Rangel"*, 49(2), 43-50.
- Clairin, R., & Brion, P. (2001). *Manual de muestreo*. Madrid: La Muralla, S.A.
- Cuestas, E. (2009). Variables. *Revista de la Facultad de Ciencias Médicas*, 66(3), 113-117.
- García, E. (2018). ¿Qué es un hábito? Acerca de la posibilidad de una desambiguación de la noción merleau-pontiana. *Ideas*, (6), 41-71.
- García, E., Gil, J. y Rodríguez, G. (2019). *Análisis factorial*. Madrid: La Muralla, S.A.
- González, S. y Carmen, M. (1997). *Percepción y conocimiento del espacio físico a lo largo del desarrollo evolutivo: socialización ambiental y educación*. La Coruña: Recuperado de https://ruc.udc.es/dspace/bitstream/handle/2183/6580/RGP_1-51.pdf?sequence=1&isAllowed=y.
- Luna, D., Manzanares, M., Rodríguez, K. y López, H. (2018). Memoria espacial a largo plazo en humanos entrenados en un laberinto virtual. *Acta colombiana de Psicología*, 21(1), 70-82. doi:<http://www.dx.doi.org/10.14718/ACP.2018.21.1.4>
- Medina, M. (2007). las pruebas de aprovechamiento estandarizadas como instrumento de medición y político. *Revista Pedagogía*, 40(1), 145-187.
- Miller, K., Willson, S., Chepp, V., & Padilla, J. (2014). *Cognitive Interviewing Methodology*. New Jersey: John Wiley & Sons.
- Morales, P., Urosa, B. y Blanco, Á. (2003). *Construcción de escalas de actitudes tipo Likert*. Madrid: La Muralla S.A.
- Morales, P., Urosa, B. y Blanco, Á. (2003). *Construcción de escalas de actitudes tipo Likert*. Madrid: La Muralla S.A.
- Municipio de Loja. (2003). *El Reciclaje Oportunidades Para Reducir la Generación de los Desechos Sólidos y Reintegrar Materiales Recuperables en el Círculo Económico*. Loja: Documento digital.
- Ramírez, J. y Gutiérrez, R. (2018). Educación y conciencia ambiental en estudiantes de dos colegios técnicos nocturnos de la provincia de Cartago, Costa Rica. *INNOVACIONES EDUCATIVAS*, 20(28), 53-65.
- Soler, S. (2013). Los constructos en las investigaciones pedagógicas: cuantificación y tratamiento estadístico. *Atenas*, 3(23), 84-101.
- Soto, S. (2018). Variables, dimensiones e indicadores en una tesis. *Tesis Ciencia*, 1-3.

Villasís, M. y Miranda, M. (2016). El protocolo de investigación IV: las variables de estudio. *Revista Alergia México*, 63(3), 303-310.

Willis, G. (2015). *Analysis of the Cognitive Interview in Questionnaire Design*. New York: Oxford University Press.