

Validación Interna de Instrumentos de Investigación en las Ciencias Sociales

Internal Validation of Research Instruments in Social Sciences

Yunkor-Romero, Yurela Kosset

Universidad Autónoma del Perú

Yurela.yunkor@autonoma.pe

Ochoa-Pachas, José Mario

Universidad Autónoma del Perú

josmar59@gmail.com

Recibido el **15.03.2021**

Aceptado el: **28.04.2021**

RESUMEN

Medir variables categóricas es para muchos jóvenes investigadores un misterio que se tiene que descubrir paulatinamente y por ello se hace necesario aclarar el panorama. Esto debido fundamentalmente a que no se puede prescindir del natural proceso de la ciencia en todo su contexto. Muchos se quedan en lo cualitativo tratando de distanciar el proceso de evolución natural de la investigación. Lo cierto está que la ciencia, su crecimiento y desarrollo son un proceso único, el cual se divide por razones didácticas más que por apreciaciones subjetivas. Un científico no puede utilizar las herramientas estadísticas para hacer estudios cuantitativos sino tiene la claridad y profundidad de los conceptos que va a manejar, desconoce las teorías que se requieren utilizar para darle respaldo a los resultados cuantitativos. Ese tránsito, que es aparentemente sencillo en las investigaciones que se dan en las ciencias naturales por tener instrumentos mecánicos, se dificulta cuando se realiza en las ciencias sociales las cuales tienen que crear y desarrollar instrumentos documentales para medir a las variables y sus dimensiones. Los autores buscan orientar a los investigadores este camino que genera una aparente controversia, pero a la luz del principio de parsimonia se esclarecerá en este artículo.

Palabras clave

Validación de instrumentos, variables categóricas, estadística.

ABSTRACT

Measuring categorical variables is for many young researchers a mystery that has to be discovered gradually and therefore it is necessary to clarify the panorama. This is mainly due to the fact that the natural process of science cannot be dispensed with in all its context. Many stay in the qualitative, trying to distance the process of natural evolution from research. The truth is that science, its growth and development are a unique process, which is divided for didactic reasons rather than for subjective appreciations. A scientist cannot use statistical tools to carry out quantitative studies if he does not have the clarity and depth of the concepts that he is going to handle, he does not know the theories that are required to be used to support quantitative results. This transit, which is apparently simple in the investigations that occur in the natural sciences because they have mechanical instruments, is difficult when it is carried out in the social sciences, which have to create and develop documentary instruments to measure the variables and their dimensions. The authors seek to guide researchers along this path that generates an apparent controversy, but in light of the principle of parsimony, it will be clarified in this article.

Keywords:

Validation of instruments, categorical variables, statistics.

SUMARIO

I. Introducción; II. Variables y su clasificación; III. Matriz y operacionalización de las variables categóricas; IV. Las dimensiones e indicadores de las variables categóricas; V. Relación entre las variables, dimensiones, indicadores, preguntas y rangos con la estadística; VI. Validación Interna: validez de contenido, validez de constructo y validez de confiabilidad; VII. Conclusiones; VIII. Recomendaciones; IX. Referencias

CONTENIDO

I. Introducción

Gran controversia se produce en la investigación cuando se trata de elaborar instrumentos de medición en las ciencias sociales debido a que el desarrollo de esta técnica se ha producido en el terreno de la psicología y otros científicos sociales preferirían tener en su campo profesional herramientas exclusivas. Lo cierto es que la confusión radica en la denominación: psicometría ya que pareciera que se circunscribe exclusivamente a la psicología; entonces también se podría crear métrica para la sociología, para la antropología, para la educación y para el derecho, utilizando las mismas herramientas, pero dentro de cada enfoque profesional. Otro factor que se debe considerar es la construcción de los instrumentos documentales puesto que algunos investigadores prefieren utilizar aquellos que se encuentran estandarizados y otros prefieren construir sus propias herramientas, aunque no cumplan con los parámetros señalados por la técnica psicométrica. Tafur & Izaguirre (2015) señalan “Si existe un instrumento validado y confiable, lo adoptará para hacer las mediciones que necesita. En caso de no existir tendrá que crear el instrumento...” (pág. 197); en ese sentido, si se busca un instrumento validado, “El investigador o el equipo de tesis puede revisar los catálogos de pruebas estandarizadas y elegir aquella que mida la variable desde el punto de vista del estudio, dándoles prioridad a la que esté adaptada a una realidad similar a la investigada” (Arbaiza Fermini, 2013, pág. 220). Además, se tiene que considerar a quien se va a encuestar y como se realizará esta operación; el primer parámetro tiene que ver con la población y el segundo a como se elige a los encuestados; lo cierto es que para un profesor, frecuentemente su población son sus estudiantes, para un sociólogo es una comunidad, para un antropólogo es una tribu, para un psicólogo es su paciente, para un abogado, que siempre apunta al experto o perito en el derecho suele encuestar justamente a los jueces, fiscales y abogados sin tomar en consideración a los justiciables; ¿por qué realiza siempre la misma operación?, es un tema que se tiene que dilucidar. Y respecto a cómo se elige a los encuestados se tiene que recordar que si la población es pequeña y accesible se puede encuestar a todos los integrantes de la población ya que el investigador estudia a la población; solo cuando la población es inalcanzable, inaccesible o muy grande se utiliza el artificio de la muestra que debe ser representativa.

II. Variables y su clasificación

Las variables en los estudios cuantitativos son las materias primas de la investigación de este enfoque. Con estos insumos se construyen el enunciado del estudio, los dilemas de investigación, los objetivos estadísticos, las hipótesis, se elaboran los instrumentos de medición, se realiza la discusión de los resultados, se plantean las conclusiones y las sugerencias, se construyen las matrices, tanto de operacionalización como la de consistencia. Supo (2015) comprendió meridianamente y planteó la importancia de las variables en los estudios de corte cuantitativo.

Se puede definir la variable como una característica, propiedad, componente de un hecho, fenómeno que se puede observar, que se puede dimensionar y que se puede medir. Estas tres características son fundamentales para poder manejar la variable y podamos comprender su operacionalización.

Las variables suelen clasificarse, por sus características intrínsecas, en categóricas y numéricas. A las variables categóricas se les suele denominar cualitativas o subjetivas; mientras que a las numéricas se les puede denominar en cuantitativas u objetivas. Los autores usan de manera indistinta estas divisiones que suelen confundir a los estudiantes ya que una variable que es cualitativa, no implica que el estudio sea cualitativo. Es necesario entender que si la variable es categórica o cualitativa o subjetiva y obtiene sus dimensiones para medirla, el estudio es cuantitativo. El solo medir la variable, es solo buscar cuantificarla implica que el estudio es cuantitativo. En los estudios cualitativos las categorías se descomponen y no se miden.

Las variables, categóricas o numéricas, presentan determinados atributos que permiten tener ciertas características que las hacen susceptibles de ser medidas de manera más precisa y exacta. Las variables categóricas pueden ser nominales y categóricas; mientras que las variables numéricas pueden ser de intervalo y de razón. Cada una de ellas se distingue de la otra por los atributos que poseen. Una variable categórica nominal solo tiene el atributo de categoría y nada más y el ejemplo más común es el sexo: masculino, femenino, no existe

orden, solo las categorías que se muestran. Por otro lado, se tiene la variable categórica ordinal, el cual se diferencia de la anterior variable por poseer el atributo orden, esto implica que existe un grado, una jerarquía entre cada componente de la variable, así, por ejemplo, el grado académico que puede ser bachiller, maestro, doctor, pos doctor y ello implica una secuencia para ser maestro se tiene que ser bachiller; para ser pos doctor primero se debe ser doctor.

Las variables numéricas, pueden ser de intervalo o de razón. Las variables de intervalo, tienen un atributo que no tienen las variables categóricas ordinales, que es la distancia, ello implica que se pueden sumar sus componentes porque cada uno tiene la misma distancia respecto del otro, como es el caso de la temperatura ya que dos grados Celsius se suman a tres grados Celsius dan como resultado cinco grados Celsius. Esto no ocurre con las variables categóricas ordinales, dos maestros no dan como suma un doctor. Lo cierto es que las variables numéricas de intervalo se encuentran relacionadas con las variables numéricas discretas que se obtienen del conteo. Por otro lado, las variables numéricas de razón, tienen un atributo que se denomina de origen, es decir que todas estas variables tienen un punto de inicio como por ejemplo el peso que se mide en una balanza (instrumento mecánico) y por obvias razones nadie va a pesar menos de 0 kilogramos; ni siquiera un electrón podría dejar de tener masa, pero esta es despreciable por ser extremadamente pequeña. Estos atributos que se van agregando a las variables de manera que la variable numérica de razón tiene los atributos origen, distancia y orden, los podemos leer en la Tabla 1.

Tabla 1
Clasificación de las Variables por sus Atributos

Variable	Clasificación	Atributo
		- Origen
	De Razón	- Distancia
		- Orden
Numérica		- Categoría

		- Distancia
	De Intervalo	- Orden
		- Categoría
		- Orden
Catagórica	Ordinal	- Categoría
	Nominal	- Categoría

Nota: Atributo, clasificación y variables.

También se puede clasificar las variables de acuerdo al nivel de investigación que se va a desarrollar, Supo (2015) ha realizado una interesante y muy clara propuesta que permite entender cómo funcionan las variables en los diversos momentos de la investigación en el enfoque cuantitativo.

En los estudios descriptivos, los cuales son univariados, la variable de estudio es denominada de interés, ya que esta variable es la que nos interesa investigar. Ello no implica que existan otras variables a las que se les denomina de caracterización (Supo, 2015), para el presente trabajo se prefiere denominarlas factores de caracterización las cuales se encuentran dentro del entorno de la variable de interés.

En los estudios relacionales, que algunos prefieren denominar correlacionales y que su característica principal es que son bivariados, es decir tienen dos variables, se les clasifica como variable supervisora y variable asociada. La variable supervisora es aquella que fue la variable de interés en el estudio descriptivo; la variable asociada se obtiene de los factores de caracterización. La variable supervisora que se denomina así porque es la variable de estudio y marca la pauta de este estudio para relacionarla con la variable asociada y determinar si tienen relación no causal y que tan fuerte es esa relación.

En los estudios explicativos, que son multivariados, siendo la mínima cantidad de variables dos, que son de tipo causal se tiene al menos una variable independiente y otra dependiente. Puede ser que la variable asociada se transforme en variable independiente siempre que exista relación con la variable de estudio y que este vínculo sea muy fuerte; la variable dependiente,

que es la variable de investigación, proviene de la variable de supervisión y es la principal variable. Se debe tener en cuenta, que solo en este tipo de investigación explicativa se busca causalidad y por ende hay variables independientes y dependientes; y son las variables independientes las que se pueden controlar y manipular, por lo que son estudios intervencionistas.

En los estudios predictivos, que también son multivariados, se tienen las variables predictoras las cuales son las que provienen de las variables independientes; y las variables a predecir que provienen de las variables dependientes y que son las variables de estudio. Estos estudios permiten predecir, pronosticar y prever los fenómenos que se pueden producir en un tiempo determinado y tomar las medidas para que los riesgos se minimicen y por ello es importante determinar de manera certera, las variables que serán predictoras y las a predecir. En estos estudios se controlan y manipulan las variables predictoras, por lo que se les clasifica como estudios de intervención.

En los estudios aplicativos, las variables pueden ser exógenas y endógenas. Las primeras provienen de las variables predictoras de los estudios predictivos; mientras que las endógenas provienen de las variables a predecir. Las variables endógenas son pues las variables de estudio. Estos estudios son de intervención y permiten desarrollar programas de mejora o de optimización. Esta clasificación se puede ver reflejada en la Tabla 2.

Tabla 2
Tipos de Variables por el Nivel de Investigación

Nivel de Investigación	Denominación de la Variable	
	Variables	Variable de Estudio
Aplicativo	Exógenas	Endógena
Predictivo	Predictoras	A predecir
Explicativo	Independientes	Dependiente
Relacional	Asociada	Supervisora
Descriptivo	Factores de caracterización	Interés
Exploratorio	Se investigan Categorías	

Nota: Denominación de las variables por su nivel de investigación.

En ese orden de ideas, se pueden vincular, las variables de los niveles de investigación con las variables de acuerdo a sus atributos. No existe ni conflicto ni contradicción. Al contrario, se observa un complemento sustantivo que permite que el investigador pueda tratar de manera adecuada la función que cumple la variable en su nivel de investigación y con qué atributos cuenta. Por ello, si el estudio es descriptivo, se puede trabajar con una variable numérica, frecuentemente utilizada en los trabajos de ciencias naturales o una variable categórica usada en los estudios sociales; y ambas variables de estudio son de interés. En un estudio relacional, también se puede trabajar con variables numéricas o categóricas y como la investigación es bivariada, una será la variable supervisora (numérica o categórica) y la otra será la variable asociada (numérica o categórica). Estas características se pueden apreciar en la Tabla 3 del presente artículo.

Tabla 3

Variables por sus Atributos y sus Niveles de Investigación

Niveles de Investigación	Denominación de la Variable de Estudio		
	Por sus Atributos		Por su ubicación en el Nivel
	Categórica	Numérica	
Aplicativo	x	x	Endógena
Predictivo	x	x	A predecir
Explicativo	x	x	Dependiente
Relacional	x	x	Supervisora
Descriptivo	x	x	Interés

Nota: Denominación de las variables por sus atributos y nivel de investigación.

III. Matriz y operacionalización de las variables categóricas

Una matriz es una tabla donde se ubican los componentes principales de un trabajo de investigación: título, problemas, objetivos, hipótesis, variables y dimensiones y la metodología. Este molde permite identificar rápidamente las probables incongruencias entre sus elementos constitutivos y alinearlos para que se pueda desarrollar de manera coherente y cohesionada. Los validadores

requieren de esta matriz para evaluar los instrumentos documentales para la investigación.

Las variables categóricas, deben ser operacionalizarlas antes de incorporarlas a la matriz de consistencia, para lo cual se debe conocer a profundidad a las variables con las que se va a trabajar. La operacionalización de las variables implica descomponerlas desde diferentes criterios o enfoques; uno puede ser en sus elementos constitutivos, otro puede ser por sus características, también se le puede descomponer por sus propiedades; si usted utiliza una teoría puede encontrar las dimensiones de la variable en función de aquella que está usando para su estudio, a falta de una teoría podría tomar en consideración una doctrina o una disciplina según sea el caso.

En las ciencias de la educación, existe mucha influencia de la psicología y por ello muchos instrumentos que se han elaborado para los estudios clínicos, es decir, para diagnosticar determinadas características se utilizan en la educación incorporando sesgos en las medidas. Lo más recomendable es que se construyan los instrumentos de acuerdo a las necesidades de la ciencia en la cual se estudia y de las variables que se están descomponiendo. Los estudios sociológicos y antropológicos frecuentemente realizan trabajos de campo cualitativo. En el campo del derecho, existe renuencia para realizar trabajos cuantitativos, incluso las variables son tratadas teóricamente y sin tener en consideración sus dimensiones e indicadores para que se puedan medir de manera estadística. La Tabla 4 puede orientar la operacionalización de las variables en estudios de las ciencias sociales.

Tabla 4

Matriz de Operacionalización de las Variables

Variable	Dimensiones
Beneficios Laborales	- Vacaciones
	- Gratificaciones
	- Compensación de Tiempo de Servicios
	- Asignación Familiar

Nota: Dimensiones de la variable beneficios laborales.

Se dimensiona la variable categórica para que se pueda medir, en este caso la variable beneficios laborales tiene cinco (5) dimensiones: vacaciones, gratificaciones, compensación de tiempo de servicios, asignación familiar y seguro de vida. Es un paso para la intención del investigador de medir la variable categórica. ¿Podría haber otras dimensiones? Si, se podría elaborar otra matriz de operacionalización de esta variable colocando otras dimensiones ya que depende del enfoque, la profesión y las teorías o disciplinas que utilice o maneje el investigador. Siendo una probable dificultad, es desde nuestra perspectiva una oportunidad de enriquecer el campo de investigación ya que para un abogado o contador esta dimensionalización de los beneficios laborales puede ser correcta, para un psicólogo podrían ser, por ejemplo, el crecimiento laboral, el reconocimiento, condiciones de trabajo.

IV. Las dimensiones e indicadores de las variables categóricas

Una vez que se ha dimensionalizado la variable categórica, se debe de establecer para cada dimensión sus correspondientes indicadores. El indicador es el parámetro para poder medir indirectamente la variable. Puede ser que, en una variable categórica se pueda pasar directamente al indicador, sin embargo, es preferible pasar por la dimensionalización para darle fortaleza y seguridad de que la variable será medida de manera correcta y confiable.

Una dimensión de cualquier variable categórica puede tener uno o más indicadores. De estos indicadores se elaboran las preguntas para la construcción del instrumento documental de investigación. El indicador tiene la peculiaridad de establecer una medida, sea discreta o continua la cual favorecerá la medida de las variables que se están investigando, tal como se lee en la Tabla 5.

Tabla 5

Matriz de Operacionalización de la Variable, Dimensiones e Indicadores

Variable	Dimensiones	Indicadores
Beneficios Laborales	- Vacaciones	- Tiempo de vacaciones
	- Gratificaciones	- Número de gratificaciones
		- Monto de cada gratificación
	- Compensación de Tiempo de Servicios	- Monto de cada depósito
		- Número de depósitos
	- Asignación Familiar	- Monto de la Asignación Familiar
	- Seguro de Vida	- Monto del Seguro de Vida
		- Número de beneficiarios

Nota: Indicadores y dimensiones de la variable beneficios laborales.

De acuerdo a lo apreciado, las dimensiones se han descompuesto en sus respectivos indicadores, y éstos permitirán elaborar las preguntas en función de las necesidades investigativas del estudioso, que por demás está señalar que no es la única cantidad de indicadores ni tampoco las únicas dimensiones. Esta riqueza, más que una debilidad es una fortaleza en las ciencias sociales.

V. Relación entre las variables, dimensiones, indicadores, preguntas y rangos con la estadística

Una vez que se ha descompuesto las variables en sus dimensiones e indicadores, se puede pasar a la elaboración de las preguntas correspondientes, las cuales de acuerdo a la técnica deben de realizarse en función del interés de lo que quiere investigar el autor. Desde el punto de vista técnica se deben tener en consideración:

- 5.1. Al menos 100 preguntas
- 5.2. Del total revisar y reducir a 70 preguntas
- 5.3. De las 70 preguntas revisar y reducir entre 30 a 40 preguntas
- 5.4. El número de preguntas puede estar entre 15 a 30 preguntas y se pueden ir reduciendo en función de la investigación.

Asimismo, la técnica indica que la mitad de las interrogantes deben ser directas y la otra porción indirecta. El objeto de este tratamiento es para buscar variabilidad en las respuestas y no haya sesgos en los mismos. Las preguntas

directas de alguna manera orientan al encuestado a un sí o un no rotundo: mientras que las preguntas indirectas hacen que el contestador reflexione y busque la respuesta que considere pertinente. Se debe tener en consideración que si la variable es categórica nominal se debe utilizar un cuestionario para cuando se quiere determinar su confiabilidad se utilice el estadístico Kuder Richardson 20 (KR-20); y cuando la variable es categórica ordinal se determina su confiabilidad utilizando el Coeficiente Alfa de Cronbach, previamente se aplica la prueba piloto correspondiente.

También se puede establecer las dimensiones de una variable utilizando la estadística, básicamente en la validación de constructo, lo que da inicio a la validación cuantitativa del instrumento de medición. Los dos caminos son viables, más si queremos estandarizar el instrumento que se está construyendo se necesita tanto la validación cualitativa como la cuantitativa utilizando las herramientas que establece la técnica psicométrica.

VI. Validación Interna: validez de contenido, validez de constructo y validez de confiabilidad

La validación se refiere a que un instrumento de investigación mida lo que debe medir y que pueda leerse la variabilidad, la confiabilidad, la estabilidad, el criterio y el rendimiento de la herramienta de medida.

De acuerdo con Puerta y Marín (2015) “La validez de contenido se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide. Un instrumento requiere tener representados a prácticamente a todos o la mayoría de los componentes del dominio del contenido de las variables a medir” (pág. 2)

En cuanto a la validación interna, se refiere a la validación suficiente y necesaria para que se pueda aplicar el instrumento a una población o a una muestra representativa de un estudio descriptivo o relacional. Esta validación no busca estandarizar el instrumento, sino hacerlo válido para un determinado

grupo de observaciones, de datos, de información que provienen de una matriz de datos.

La matriz de datos es aquella que se mide en una sola oportunidad o en más y no se modifica, es como una fotografía ya que los datos siempre son los mismos una vez recogidos de la muestra representativa o de la población, el tiempo no modifica los valores de la matriz de datos; es distinta a una base de datos, la cual está en permanente cambio y va modificándose en el tiempo, la cual se puede ver en la variación permanente de la cuenta bancaria de la población de Lima Metropolitana, o los pagos de impuestos a la Sunat, o el número de nacimientos que suceden en un determinado día. La matriz de datos se procesa, ya sea de forma manual o utilizando Excel o programas estadísticos como el SPSS, el Statgraphics, el Minitab, entre otros.

Dentro de la validación interna se tienen en consideración tres: validación de contenido, validación de constructo y la validación de confiabilidad. De las tres, la que se suele excluir es la validación de constructo, ya que con esta se puede dimensionar la variable, se puede establecer la variabilidad de las preguntas, se establece la correlación entre las preguntas y entre las preguntas y el instrumento de investigación. Siendo un trabajo eminentemente estadístico muchos prefieren obviarlo, sin embargo, es necesario realizarlo si se quiere estandarizar el instrumento de medida.

6.1. Validación de Contenido

La validación de contenido es el primer peldaño para validar una herramienta de medida en el campo investigativo, en ese sentido es cualitativa, donde los expertos evalúan las preguntas que el investigador ha diseñado para la recopilación de sus datos utilizando las técnicas de la observación, documentación y la entrevista. En ese sentido, Denzil & Lincoln (2000) declaran que “La investigación cualitativa es una actividad que sitúa al observador en el mundo...y consiste en una serie de prácticas interpretativas que hacen el mundo visible. Estas prácticas interpretativas transforman el mundo, pues lo plasman en una serie de representaciones

textuales a partir de los datos recogidos en el campo mediante observaciones, entrevistas, conversaciones, fotografías, etc.” (pág. 3). Por ello el investigador debe interpretar la realidad que busca estudiar.

Ha existido una gran controversia respecto a la validación de contenido y en ese sentido, Escurra (1988) expresa “Por lo general esta toma de validez se ha determinado mediante la comparación sistemática de los ítem de la prueba con el dominio de contenido estudiado, este análisis es factible de ser llevado a cabo de dos formas, la primera que consiste en estudiar de manera lógica y racional los ítem explicitando por qué se incluye en la prueba; y la segunda en la cual con ayuda de un grupo de jueces competentes y calificados se evalúa el grado en que los reactivos concuerdan con los planteamientos del constructo con el instrumento, siendo denominada esta técnica como el criterio de jueces (Andreani 1975); y que en muchos casos ha sido la estrategia usada por excelencia para evaluar la validez de contenido (Aiken, 1980)” (pág. 105).

Para realizar la validación de contenido, lo primero que se hace es revisar la literatura existente actualizada para determinar si hay un instrumento que mida las variables que se están estudiando; se debe determinar si estos instrumentos se han estandarizado y si están en idioma inglés y/o español; además si la versión que se ha elegido está calibrada o actualizada.

Por otra parte, se debe identificar si el concepto o los conceptos que se van a investigar se encuentran plenamente definidos, parcialmente definidos o no se encuentran definidos. Cuando el concepto está definido en toda su extensión se establece que se sustenta en una teoría muy consistente. En ese sentido, los textos y los artículos de investigación son nuestras fuentes que permiten que desarrollemos el instrumento de medición o aplicar una prueba estandarizada es decir validada tanto interna como externamente.

Si el concepto está definido parcialmente, es decir existen muchas definiciones y por ende teorías que explican el concepto, entonces debemos explorar y decidir qué teoría, aunque no esté terminada, permitirá sustentar

nuestro trabajo de investigación ya que se debe entender que la decisión que tome el investigador debe tener en consideración dos aspectos: que tenga claro lo que va a investigar y que debe elaborar el instrumento que permita medir aquello. Si el concepto carece de definición, entonces el investigador debe construir el concepto a través de la exploración del mismo, lo que permite delimitar nuestro campo de estudio.

Cuando el concepto está parcial o no está definido se debe realizar una entrevista a profundidad a los expertos –aquellas personas que conocen el tema de estudio- con el objetivo de tener la materia prima para construir el concepto que necesitamos medir en nuestra investigación. Se debe recordar que la entrevista a profundidad se inicia con una sola pregunta y para evitar la tensión con el entrevistado, no tiene libreto, no tiene una estructura y por ende se orienta hacia una conversación informal, cualitativa, sin reglas, holística, que permite identificar las percepciones del entrevistado con el propósito de encontrar las características comunes y relevantes del concepto. Esto implica que debemos de adaptarnos a los entrevistados y tener, a través de un cuaderno de notas, de una grabadora o de un cámara de video, un registro de la entrevista que estamos realizando.

Lo que se obtiene de la entrevista a profundidad realizada a los expertos, son las palabras clave, las cuales ayudarán a construir el concepto. Estas palabras deben ser registradas en un cuaderno de registro sin tener ningún tipo de criterio o convención, aun cuando estén repetidas estas palabras. Lo que sí se puede utilizar como un criterio orientador, es que si se quiere tener un instrumento que tenga 25 preguntas, se deben tener al menos cinco veces ese número, en este caso, 125 palabras clave.

Una vez que se tengan las palabras clave, éstas se agrupan y si el concepto se encuentra parcialmente definido se pueden utilizar las dimensiones que el investigador ha aceptado como parte de su trabajo. Si existieran palabras que no tienen ubicación, entonces se coloca en otro grupo y teniendo en mente que el investigador puede plantear nuevos términos y nuevos conceptos. Este procedimiento se realiza para reducir a la mitad el número

de palabras clave, ya que al clasificarlas por grupos o dimensiones se van descartando los términos repetidos y los términos ambiguos, por lo que de 125 quedarían 62 palabras clave.

El siguiente paso es la formulación de los reactivos del instrumento de investigación, ello implica que como quedaron 62 palabras clave, entonces se deben formular 62 reactivos. En este punto se debe realizar una entrevista enfocada en las personas que conforman la población de estudio, tomando como elementos de la entrevista los 62 reactivos que se han construido con el propósito de establecer si la población de estudio identifica estas características. Producto de esta entrevista enfocada podemos ampliar el número de reactivos, reducirlo y establecer si se va a construir un cuestionario, una escala o un inventario. En el caso que se quiera evaluar el nivel de conocimientos, se utiliza un cuestionario; si se va a evaluar actitudes, conductas u opiniones se utiliza una escala; si se van a evaluar ambos, se puede utilizar un inventario.

Si se construyen cuestionarios, que deben tener entre cuatro o cinco alternativas de respuestas, con preguntas cerradas deben evitarse alternativas como todas las anteriores o ninguna de las anteriores; asimismo tampoco debe combinarse las respuestas correctas con varias alternativas. En cambio, si construyes escalas, las alternativas deben ser graduadas, es decir, desde nunca, a veces, casi siempre, y siempre.

Después de tener, el cuestionario o la escala con los 62 reactivos, sigue el juicio de jueces ya que los expertos ayudaron en la entrevista a profundidad y la población de estudio ayudó en la entrevista enfocada. El juez conoce la metodología de la investigación y no necesariamente tiene conocimiento del tema de investigación; esto debido a que el juez ayuda a evaluar los ítems que se han formulado, por lo mismo que conocen y manejan la metodología correspondiente.

Al tener un listado de 62 preguntas, estas se someten a la evaluación de los jueces, para lo cual se debe de hacer una selección correcta y adecuada de

los mismos. Lo aplicado por convención es elegir a cinco jueces de diversos campos del conocimiento para evitar sesgos a la hora de evaluar las preguntas de investigación. La función del juez es revisar los reactivos en función de la suficiencia, la pertinencia y la claridad con que han sido redactados.

Un reactivo es suficiente cuando cubre todos sus componentes y características; un reactivo es pertinente cuando no excede ni sus componentes ni sus características; y un reactivo es claro cuando la persona que lo lee lo entiende y sabe a qué se refiere el concepto, por lo que se debe evitar utilizar términos técnicos que puedan confundir a la persona encuestada.

La evaluación por jueces permitirá reducir el número de reactivos, en un 20% (criterio convencional), es decir, de los 62 reactivos, se tendrán 50. Se debe recordar que los jueces realizan una evaluación cualitativa y no se utilizan procedimientos estadísticos, ya que los jueces sugieren si los reactivos son suficientes, relevantes, convenientes y transparentes. En este punto, se tiene el instrumento con 50 reactivos y que tiene validez de contenido, para pasar a la siguiente fase.

Se tiene que comprender que la validación de contenido es netamente cualitativa no se utiliza estadística, sin embargo Ecurra señala (1988) “El presente estudio evalúa tres formas de cuantificar la validez de Contenido por criterio de Jueces: el Índice de Acuerdo (IA), la Prueba Binomial (PB) y el coeficiente V de Aiken (V); computados en base a todas las respuestas posibles asignadas por 10 jueces a un ítem. Los resultados permiten concluir que el coeficiente V de Aiken es el más adecuado para determinar este tipo de validez, ya que permite obtener valores factibles de ser contrastados estadísticamente según el tamaño de la muestra de jueces seleccionada” (pág. 103). Esta afirmación que es necesaria incorporarla para información de los lectores, contradice el propio concepto de validación de contenido que es sustantivamente teórica.

6.2. Validación de Constructo

Una vez culminada la validación de contenido, se empiezan a medir las propiedades métricas del instrumento que se ha construido, por lo que con la validación de constructo se inicia la etapa cuantitativa de la validación que permite evaluar las propiedades métricas del instrumento.

Se formularán los reactivos en dos aspectos, la mitad deben ser directos o favorables –cuando son afirmativos respecto al concepto- y la otra mitad deben ser indirectos o desfavorables –cuando son negativos respecto al concepto; se deben tener en cuenta que para formular los reactivos indirectos se debe evitar utilizar las palabras no, nunca. La puntuación de los reactivos directos debe marcar la afirmación como la más alta; mientras que, para los reactivos indirectos, la puntuación debe ser invertida. Los reactivos se redactan en dos sentidos para establecer variabilidad en las respuestas ya que sin ellas no podríamos medir sus propiedades métricas.

El orden de los reactivos debe ser aleatorio, para reducir los sesgos y la subjetividad al aplicarlo a una población similar a la que se va a estudiar, esta prueba se denomina Piloto, y que debe ser aplicada por el investigador ya que nuevamente se va a evaluar la claridad de los reactivos, es decir si están redactados de manera correcta y no genera dudas o incertidumbres entre los encuestados en la prueba piloto. Para Muñiz & Fonseca-Pedrero (2019) “La finalidad de cualquier estudio piloto es examinar el funcionamiento general del instrumento de medida en una muestra de participantes con características semejantes a la población objeto de interés. Esta fase es de suma importancia ya que permite detectar, evitar y corregir posibles errores, así como llevar a cabo una primera comprobación del funcionamiento del test en el contexto aplicado. El estudio piloto podría verse como una representación en miniatura de lo que posteriormente va a ser el estudio de campo” (pág. 11), de ahí la importancia del mencionado estudio.

La aplicación del instrumento debe tener en consideración diversos factores y en ese sentido, se coincide con Muñiz & Fonseca-Pedrero (2019) “Las actividades relacionadas con la aplicación y el uso del instrumento de medida son cruciales durante el proceso de validación (Muñiz y Bartram, 2007; Muñiz et al., 2005). Cuando aplicamos cualquier instrumento de medida hay que cuidarse de que las condiciones físicas de la aplicación sean las adecuadas (luz, temperatura, ruido, comodidad de los asientos, etc.). Igualmente, las personas encargadas de la administración del instrumento de medida deben establecer una buena relación (*rapport*) con los participantes, estar familiarizados con la administración de este tipo de herramientas, dar las instrucciones a los participantes correctamente, ejemplificar con claridad cómo se resuelven las preguntas, supervisar la administración y minimizar al máximo las posibles fuentes de error. Un aspecto que no se debe descuidar es el referido a las instrucciones en la entrega del material. Las directrices utilizadas (el qué dice y el cómo lo dice) por evaluador en la administración del test pueden modificar drásticamente las respuestas de las personas” (pág. 12).

Alcanzados los resultados del ensayo piloto, se reducen los reactivos mediante procedimientos estadísticos, que en este caso se tienen 25 reactivos directos y 25 reactivos indirectos. En ese sentido, si se evalúa la opinión de la población de la prueba piloto de un determinado concepto, los puntajes más altos indican que la opinión es favorable o positiva; y si los puntajes son bajísimos, señalan que la opinión es desfavorable o negativa. Siendo el orden lógico, los puntajes más altos se obtendrán de la suma total y del mismo modo en la puntuación más bajo. Estos resultados se pueden verificar utilizando el índice de correlación (Pearson para variables numéricas; y Spearman para variables categóricas ordinales). Por ello, un reactivo es consistente si la correlación es positiva entre cada reactivo y la suma total correspondiente. Cuanto más altos sean los índices de correlación, mejor ubicados estarán en la suma total por lo que tendrán una importante participación en el resultado final del puntaje del instrumento que estamos evaluando.

Tanto para la de correlación de Pearson o Spearman, si los valores de correlación entre el reactivo y el total de valores son superiores a 0,8 señalan una muy buena participación del reactivo en el resultado total, ya que los valores de correlación fluctúan entre 0 y 1. A este procedimiento donde se correlaciona el ítem con la suma total se le denomina correlación ítem-total; y para que exista una excelente correlación debe haber una buena dispersión de los resultados, tanto en los reactivos como en la suma total de estos. Asimismo, como la dispersión o variabilidad se mide a través de la varianza, se deben obtener altos valores de la misma en cada reactivo y en la suma total de los mismos.

Establecidos estos criterios, se debe reducir los reactivos de 50 a 30, es decir eliminar 20 reactivos. Estos reactivos se eliminan ordenando los mismos según el índice de correlación y el resultado total, de mayor a menor de tal manera que los últimos 20 reactivos tienen la menor magnitud de correlación respecto al puntaje total. Como se ha establecido el índice de correlación tiene que ver con la variabilidad y en ese sentido, cuando un reactivo no expresa dispersión en las respuestas es porque los encuestados han respondido en el mismo sentido, entonces no habría diferencias y ello implicaría que no tendría sentido ni la construcción del instrumento ni la propia investigación.

Una vez que se tienen los 30 reactivos se tienen que agrupar en dimensiones y a este procedimiento se le denomina reducción de dimensiones, es decir los 30 reactivos se agrupan en tres, cuatro, cinco, seis dimensiones por tener características semejantes o similares. Por ello, un instrumento tiene varias dimensiones y cada una de ellas tiene uno o varios reactivos; en ese sentido al agrupar los reactivos en dimensiones, es necesario establecer si los puntajes de cada dimensión son consistentes con el puntaje final o total y si el puntaje de cada reactivo es consistente con el puntaje de cada dimensión.

Para la reducción de las dimensiones, que es parte de la validación de constructo, se realiza con el análisis factorial, el cual tiene dos etapas: el

análisis factorial exploratorio y el análisis factorial confirmatorio. El análisis exploratorio sugiere como se deben agrupar los reactivos para lo cual se deben de identificar cada grupo con una denominación, la clasificación realizada empíricamente puede ayudar para este procedimiento. La correlación obtenida sugiere dos cosas: que los reactivos se encuentran vinculados de forma numérica y que esta asociación es también conceptual. El análisis factorial confirmatorio permite dar el siguiente paso en la agrupación de los reactivos, por lo que este análisis permite corroborar las dimensiones de los reactivos para el instrumento de investigación.

6.3. Validación de Confiabilidad

Una vez establecida la validez de constructo, se debe realizar la validación de confiabilidad o de fiabilidad. La confiabilidad "...se define como el grado en que un instrumento construido por varios ítems presenta una alta correlación y miden consistentemente una muestra. Es decir, la dimensión considerada para el diseño del instrumento, sin embargo, debe recordarse que la confiabilidad no es criterio suficiente para determinar la validez de un instrumento" (Soriano, 2014, págs. 31-32). Para el Coeficiente Kuder-Richardson (KR-20) que se utilizan para los instrumentos denominados cuestionarios que tienen variables categóricas nominales dicotómicas, de acuerdo a la Tabla de los valores de dicho coeficiente:

Tabla 6

Escala de Kuder-Richardson 20 (KR-20)

Rango	Magnitud
0,81 – 1,00	Muy Alta
0,61 – 0,80	Alta
0,41 – 0,60	Moderada
0,21 – 0,40	Baja
0,01 – 0,20	Muy Baja

Nota: Thorndike (1989), Magnusson (1982).

Estos estadísticos se utilizan para el ítem-total como para las dimensiones-total y para el ítem-dimensión. Los valores permitidos varían de 0,01 a 1,00 y para que el instrumento sea validado el resultado debe ser mayor a 0,70. En la escala de Kuder-Richardson presentada, se indica que el valor aceptable está a partir de 0,61. Para los instrumentos de medición a escala, que utilizan Likert, se acepta para validar instrumentos valores mayores a 0,8 que corresponden al coeficiente alfa de cronbach, tal como se puede observar en la Tabla de los valores que corresponden a dicho coeficiente.

Tabla 7

Escala del Coeficiente Alfa de Cronbach

Valores	Magnitud
> 0,9	Excelente
> 0,8	Bueno
> 0,7	Aceptable
> 0,6	Cuestionable
> 0,5	Pobre
< 0,5	Inaceptable

Nota: George & Mallery (2003).

En otros casos, “El valor mínimo aceptable para el coeficiente alfa de Cronbach es 0.70; por debajo de ese valor la consistencia interna de la escala utilizada es baja. Por su parte, el valor máximo esperado es 0.90; por encima de este valor se considera que hay redundancia o duplicación. Varios ítems están midiendo exactamente el mismo elemento de un constructo; por lo tanto, los ítems redundantes deben eliminarse. Usualmente se prefieren valores de alfa entre 0.80 y 0.90” (Oviedo & Campos, 2005, pág. 577).

VII. Conclusiones

Primera:

Se establece que, para la elaboración de un instrumento de medición documental en la investigación, se deben seguir procedimientos específicos que se encuentran establecidos en la psicometría.

Segunda:

Existe una validación interna que está conformada por la validación de contenido que es sustantivamente cualitativa; y la validación de constructo y la validación de confiabilidad, que son fundamentalmente estadísticas.

Tercera:

Estas validaciones internas (contenido, constructo y confiabilidad) son necesarias y suficientes para el desarrollo de una investigación cuantitativa, ya que su construcción se sustenta en los criterios cualitativos y estadísticos. Si se quisiera estandarizar el instrumento, se debe de realizar la validación externa.

Cuarta:

La validación de contenido tiene la intervención de expertos y jueces; la validación de criterio utiliza la estadística para la reducción de dimensiones utilizando el análisis factorial; y la validación de confiabilidad utiliza el KR-20 para cuestionarios y el coeficiente alfa de cronbach para las escalas.

VIII. Recomendaciones

- a. Se recomienda la producción de eventos de formación de validadores de instrumentos de investigación para las ciencias sociales.
- b. Se sugiere que se establezca que los expertos son aquellas personas que conocen el concepto que se va a estudiar; mientras que los jueces son aquellos que manejan la metodología de la investigación. Frecuentemente los expertos no son jueces y los jueces no son expertos.

- c. Se sugiere tener un número impar de jueces para que puedan evaluar la suficiencia, pertinencia y claridad de los reactivos.

IX. Referencias

- Arbaiza Fermini, L. (2013). *Cómo elaborar una tesis de grado*. ESAN.
- Denzin, N., & Lincoln, Y. (2000). The discipline and practice of qualitative research. En N. Denzin, & Y. Lincoln, *Handbook of Qualitative Research* (págs. 1-28). Sage Publications.
- Escurra, M. (1988). Cuantificación de la validez de contenido por criterio de jueces. *Revista de Psicología*, 6(1-2), 105.
- Geroge, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference 11.0 update*, (4th ed.). Allyn& Bacon.
- Magnusson, D. (1982). *Teoría de los Tests*. Trillas.
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16.
- Oviedo, H., & Campos, A. (2005). Aproximación al uso del Coeficiente de Alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 577.
- Puerta, L., & Marín, M. (2015). Análisis de validez de contenido de un instrumento de transferencia de Tecnología Universidad - Industria de Baja California, México. *XX Congreso Internacional de Contduría, Administración e Informática*. México.
- Soriano, A. (2014). Diseño y validación de instrumentos. *Diálogos*, 31-32.
- Supo, J. (2015). *Cómo empezar una tesis*. Bioestadístico EIRL.
- Tafúr Portilla, R., & Izaguirre Sotomayor, M. (2015). *Cómo hacer un proyecto de investigación*. Alfaomega Colombiana.
- Thorndike, R. (1989). *Psicometría Aplicada*. Limusa.