



<https://doi.org/10.22402/ed.leed.978.607.26779.13c03>

La Metodología en el Quehacer del Psicólogo Organizacional

Methodology in the Practice of the Organizational Psychologist

Carlos Nava Quiroz y Cynthia Zaira Vega Valero

Resumen

El propósito del presente capítulo es exponer algunas de las técnicas metodológicas que pueden ser usadas, por el psicólogo, dentro de las organizaciones, así como los principios psicométricos más importantes para poder evaluar el comportamiento. Más que intentar replicar o resumir las metodologías (de grupo, encuestas y psicometría) la exposición pone su atención en algunos de los aspectos más importantes y problemáticos para los interesados en el tema y a partir de ejemplos prácticos. Para la metodología de grupos el acento está en la aleatoriedad, para homogeneizar los grupos de tratamiento, en cuanto a variables extrañas y las ventajas que presenta ésta en cuanto a la correcta elección del diseño y los alcances en la interpretación de los resultados. Para la investigación por encuesta, la aleatoriedad tiene un papel diferente, es puesta como el eje para generalizar los hallazgos encontrados en una muestra a la población. El cálculo de la muestra y el método de recolección aleatorio son importantes para poder considerar con un margen de error mínimo que lo encontrado en la muestra existe en la población bajo estudio. Aquí lo que se quiere conocer son los parámetros de la población, a partir de muestras. En el caso de los instrumentos de medida, la pretensión está puesta en que los datos derivados del instrumento sean válidos, lo que estará garantizado siempre y cuando se cumplan los criterios metodológicos de construcción del mismo, lo que involucra cierto grado de estadística. El correcto uso de la metodología tiene que ver con una buena formación del psicólogo y será en beneficio de la psicología.

Palabras clave: Selección de la muestra, metodología científica, selección aleatoria; fuentes de validez, consistencia interna

Abstract

The purpose of this chapter is to present some of the methodological techniques that can be used by psychologists within organizations, as well as the most important psychometric principles for evaluating behavior. Rather than attempting to replicate or summarize the methodologies (group, surveys, and psychometrics), the discussion focuses on some of the most significant and problematic aspects for those interested in the topic, using practical examples. For group methodology, the emphasis is on randomness to homogenize treatment groups concerning extraneous variables and the advantages this offers in terms of the correct choice of design and the scope of interpreting the results. In survey research, randomness plays a different role; it is central to generalizing the findings from a sample to the population. The calculation of the sample and the method of random collection are important to consider with a minimal margin of error that what is found in the sample exists in the population under study. Here, the aim is to know the parameters of the population from samples. In the case of measurement instruments, the intention is that the data derived from the instrument are valid, which will be guaranteed as long as the methodological criteria for its construction are met, which involves a certain degree of statistics. The correct use of methodology is related to the psychologist's good training and will benefit psychology.

Keywords: Sample selection, scientific methodology, random selection; sources of validity, internal consistency

Contenido Temático

- Resumen, 50
- Abstract, 50
- Introducción, 51
- Metodología de Grupos, 52
- Metodología por Encuesta, 55
- Instrumentos de medida, 58
- Comentarios, 62
- Referencias, 63

“Medir es conocer.”

Lord Kelvini

Introducción

El propósito del presente capítulo es exponer – de manera sucinta- algunas de las técnicas metodológicas que pueden ser usadas o implementadas, por el psicólogo, dentro de las organizaciones (dicho en el sentido más amplio del término), así como los principios psicométricos más importantes para poder evaluar el comportamiento. La idea surge como una manera de apoyar la inserción del psicólogo en las organizaciones a partir del trabajo sistemático y ordenado que la metodología ofrece.

La exposición será sumaria, poniendo atención en algunos de los aspectos más relevantes, que son problemáticos -sobre todo- para los iniciados en el tema. Es conveniente señalar que existen suficientes textos, en extenso, que abordan la metodología, en este caso recomendamos -de inicio- los escritos originales y a la postre tratados más exhaustivos. Creemos que no hay por que re escribir lo ya hecho, sino llamar la atención en aspectos que pueden ser importantes al momento de tomar decisiones metodológicas.

El psicólogo tiene que caracterizarse por ser un analista del comportamiento humano, no sólo en el ámbito de investigación sino también en lo aplicado. El psicólogo organizacional (PO) debe ostentar una identidad propia que sea característica de su hacer, para que esto ocurra tiene que poseer un bagaje teórico y una formación metodológica sólida.

El PO, por costumbre o tradición, se ha dedicado a resolver problemas inmediatos, o que en muchas ocasiones no le corresponden, como por ejemplo la contratación y liquidación de personal. Las organizaciones están constituidas por individuos que se comportan, por lo que la psicología y el psicólogo pueden ofrecer aportaciones invaluable a estas.

Una de las áreas en la que tiene mayor injerencia el PO es la selección de personal, sin embargo, su hacer puede ser enmarcado de manera más amplia. La evaluación de la satisfacción laboral, implementación de sistemas de recompensas o estímulos laborales, prevención de accidentes de trabajo, salud laboral, entre otros. En todos los anteriores la metodología juega un papel muy importante. El tipo de metodología (diseños de grupos, la investigación por encuestas y la construcción y evaluación de instrumentos de medida, entre otras) que puede ser más útil al psicólogo organizacional es aquella que está relacionada, de manera directa, con los estudios de campo, ya sea para averiguar posibles problemas dentro de la organización o la toma de decisiones.

En lo que respecta a la formación metodológica, el psicólogo sabe que no solo existe la metodología de laboratorio, (por ejemplo, con los diseños de $N=1$) y que en el trabajo de campo la metodología de grupos, encuestas, lo mismo que poseer un conocimiento teórico práctico de la medición y evaluación, entre otras, pueden ser útiles para colaborar en las organizaciones. Lo que estamos proponiendo es que el practicante de la psicología trabaje con una visión diferente a sólo resolver problemas inmediatos y que en muchas ocasiones no muestran la capacidad y el perfil real del psicólogo y la psicología. Lo que a continuación presentamos son algunos de los aspectos más importantes en la formación del psicólogo organizacional, no recurriremos a una exposición tradicional de la metodología ya que existen excelentes textos que pueden cubrir este aspecto, en todo caso nos referiremos a las habilidades específicas que es deseable tenga el psicólogo organizacional, desde el punto de vista metodológico. Este escrito es producto, en gran parte, del contacto con estudiantes de posgrado, de sus inquietudes, dudas y reflexiones.

Metodología de Grupos

Existen diferentes clasificaciones acerca del tipo de estudios en ciencia, (Kerlinger y Lee, 2002), pero el interés de este trabajo está puesto en los estudios de campo, bajo el supuesto de que no sólo los psicólogos organizacionales sino otros tantos trabajan en situaciones naturales o sociales.

Los estudios de campo son llevados a cabo en ambientes naturales; las escuelas, hospitales, empresas públicas o privadas, entre otras y el control de las variables extrañas es muy limitado. Dependiendo de los propósitos, los estudios de campo pueden ser de investigación o aplicación, bajo esta circunstancia es la metodología que se emplea (por ejemplo, de grupos, de correlación o de encuestas). Pongamos por caso que un joven PO, quiere evaluar la pertinencia del método tradicional para enseñar a leer y escribir -a estudiantes de educación básica- y contrastarlo con un método digital. Lo que hace el joven PO es construir un diseño de grupos, donde los participantes son elegidos al azar lo mismo que el grupo en el cual van a participar. En seguida, ambos grupos son sometidos a un proceso de evaluación inicial (Pretest), -que consiste en una prueba de lectura y toma de dictado - luego se introducen las variables de tratamiento y al final las medidas de prueba (postest) -prueba de lectura y dictado- que muestren si es que existieron diferencias en el aprendizaje de la lectura-escritura, respecto del método. En este caso, el propósito primario es de investigación y luego la posible aplicación, ya sea para substituir o conservar la estrategia de enseñanza.

Otro escenario es cuando -suponiendo el mismo planteamiento- no es posible asignar los participantes de manera aleatoria, los grupos ya están formados y no hay manera de modificarlos. Así, las preguntas que surgen son ¿los resultados serán los mismos que cuando aleatorizamos? ¿Qué tanta confianza puedo tener en estos resultados? La respuestas a estas interrogantes está en el conocimiento metodológico.

El seminal texto de Campbell y Stanley (1995), "Diseños experimentales y cuasiexperimentales en la investigación social", tratan sobre las fuentes de invalidez de los diseños del grupo. En primer lugar refieren las fuentes de invalidez interna y externa. Éstas fuentes de invalidez junto con la asignación al azar de los participantes o individuos a los grupos, son la columna vertebral de los diseños de grupo y su validez en general.

Así, Campbell y Stanley (et.al., pag. 16) afirman que “Llamamos validez interna a la mínima imprescindible, sin la cual es imposible interpretar el modelo”. Esto es, son variables extrañas que afectan la veracidad de la intervención en cuestión. Mientras que “... la validez externa plantea el interrogante de la posibilidad de generalización: ¿a qué poblaciones, situaciones, variables de tratamiento y variables de medición puede generalizarse este efecto?”. Por otra parte, los mismos autores declaran que “... la asignación aleatoria a los diferentes grupos de tratamiento... sirve para lograr, dentro de los límites estadísticos conocidos, la igualdad de los grupos antes del tratamiento” (pag. 18). Lo que garantiza la aleatoriedad es que los grupos sean homogéneos de inicio, así, sería de esperar que ambos fueran iguales en cuanto a diferentes variables como número de participantes, edad, género, inteligencia, etcétera.

A continuación presentamos las definiciones conceptuales de las diferentes fuentes de invalidez según Campbell y Stanley (et.al. pag. 17-18):

Fuentes de invalidez interna:

- a. Historia, los acontecimientos específicos ocurridos entre la primera y la segunda medición, además de la variable experimental.
- b. Maduración, procesos internos de los participantes, que operan como resultado del mero paso del tiempo (no son peculiares a los acontecimientos en cuestión), y que incluyen el aumento del edad, el hambre, el cansancio y similares procesos internos de los participantes, que operan como resultado del mero paso del tiempo (no son peculiares a los acontecimientos en cuestión), y que incluyen el aumento del edad, el hambre, el cansancio y similares.
- c. Administración de Test, el influjo que la administración de un test ejerce sobre los resultados de otro posterior.
- d. Instrumentación, los cambios en los instrumentos de medición o en los observadores o calificadores participantes que pueden producir variaciones en las mediciones que se obtengan.
- e. Regresión estadística, opera allí donde se han seleccionado los grupos sobre la base de los puntajes extremos.
- f. Sesgos resultantes en una selección diferencial de participantes para los grupos de comparación.
- g. Mortalidad experimental, o diferencia en la pérdida de participantes de los grupos de comparación.
- h. Interacción entre la selección y la maduración, etc...

Fuentes de invalidez externa:

1. El efecto reactivo o de interacción de las pruebas, cuando un pre-test podría aumentar o disminuir la sensibilidad o la calidad de la reacción del participante a la variable experimental, haciendo que los resultados obtenidos para una población con pretest no fueran representativos de los efectos de la variable experimental para el conjunto sin pretest del cual se seleccionaron los participantes experimentales.
2. Los efectos de interacción de los sesgos de selección y la variable experimental.
3. Efectos reactivos de los dispositivos experimentales, que impedirían hacer extensivo el efecto de la variable experimental a las personas expuestas a ella

en una situación no experimental.

4. Interferencias de los tratamientos múltiples, que pueden producirse cuando se apliquen tratamientos múltiples a los mismos participantes, pues suelen persistir los efectos de tratamientos anteriores.”

La tabla número 3.1 muestra seis diseños, los tres primeros pre-experimentales y del cuatro al seis experimentales. La X representa la intervención y la O la observación o prueba (test). El signo (-) significa que es un factor no controlado, el símbolo (+) quiere decir que la fuente de invalidez está controlada y los de interrogación que no es clara su participación en las fuentes de validez. La R representa la asignación al azar de los participantes.

De regreso a nuestro ejemplo, el grupo control es aquel que es tratado con el método tradicional de enseñanza de la lectura y escritura y el grupo experimental es el digital. La diferencia entre tener los grupos (control y experimental) y que son asignados al azar y con grupos ya formados de antemano es trascendente, si bien se conservan los niveles comparativos que los grupos ofrecen, la falta de asignación al azar hace que no se pueda garantizar la validez interna del diseño. Por lo que es muy difícil afirmar que los resultados obtenidos a partir de grupos ya formados puedan ofrecer datos con validez aceptable.

Por otra parte, hay que destacar que si bien el grupo control pretest-posttest (ver Tabla 3.1 diseño 4) cubre de manera satisfactoria las fuentes de invalidez interna el diseño 6 con sólo post-test cubre por igual todas las fuentes de invalidez interna además de la interacción en la administración de test y el tratamiento, validez externa.

			Diseños preexperimentales			Diseños experimentales		
			1.- Estudio de caso con una medición	2.- Pre-test-posttest de un solo grupo	3.- Comparación con un grupo estático	4.- Grupo control pre-test-post-test	5.- Cuatro grupos de Solomon	6.- Grupo de control con post-test únicamente
			X O	O X O	X O O	R O X O R O O	R O X O R O O R X O R O	R X O R O
Fuentes de Invalidez	Validez Interna	Historia	-	-	+	+	+	+
		Maduración	-	-	¿?	+	+	+
		Administración de Test		-	+	+	+	+
		Instrumentación		-	+	+	+	+
		Regresión		¿?	+	+	+	+
		Selección	-	+	-	+	+	+
		Mortalidad	-	+	-	+	+	+
		Interacción		-	-	+	+	+
	Validez Externa	Interacción de administración de test y X		-		-	+	+
		Interacción de selección y X	-	-	-	¿?	¿?	¿?
		Dispositivos reactivos		¿?		¿?	¿?	¿?
		Interferencia de X múltiples						

Tabla 3.1 Diseños pre-experimentales y experimentales

En este caso, es más pertinente desarrollar un experimento con dos grupos elegidos al azar y sin pretest, lo cual es más económico en términos de tiempo y representa igual nivel de confianza que el diseño 4. Respecto de este diseño 6, es común pensar o tener dudas sobre su pertinencia, en tal caso Campbell y Stanley (et.al.) señalan.

"Por razones psicológicas, es difícil renunciar a <<tener la seguridad>> de que los grupos experimentales y de control eran <<iguales>> antes del tratamiento experimental diferencial. No obstante, la aleatorización implica la mayor seguridad, aplicable a cualquier fin, de la carencia de sesgos iniciales entre grupos. Dentro de los márgenes de confianza establecidos por las pruebas de significación, la aleatorización puede ser suficiente, sin necesidad de recurrir al pretest." (Pág. 54).

En la práctica es común encontrar diseños de pretest-postest -sin asignación aleatoria- (diseños de pretest-postest de un solo grupo) y quienes lo usan ponen demasiada fe en ellos, cuando sería incluso más práctico emplear un diseño de comparación con un grupo estático (diseño 3). En resumen, el diseño 3, que es igual al 6 -solo que sin asignación al azar, presenta más ventajas que el diseño 2 de pretest-postest. Hay poco que decir respecto del diseño uno -estudio de caso con una medición- no ofrece ninguna seguridad de los efectos del tratamiento, por más que se le pregunte a los individuos o participantes acerca de su satisfacción con la intervención.

En cuanto al diseño 5 (4 grupos de Solomon) es comparable con el diseño 6, y ofrece como ventaja la posibilidad de hacer comparaciones muy específicas que brindan más información, pero también presenta el problema de los niveles de comparación, en términos de las pruebas estadísticas a ser usadas, es más costoso en términos de tiempo y participantes para asignar a los grupos.

Así, en el caso del ejemplo, el joven investigador pudo haber desarrollado un diseño aleatorizado de dos grupos con pos test y supongamos que en efecto después de aplicar una prueba estadística encuentra diferencias significativas, donde el método digital fue superior, por lo tanto; en primer lugar, hay la confianza acerca de la validez interna del diseño -es decir, las fuentes de invalidez interna fueron controladas- con lo cual es plausible concluir que el método digital es superior al tradicional para el aprendizaje de la lectura y escritura.

Sin embargo, no es posible generalizar los hallazgos en razón de tres elementos, el primero de ellos está en que el propósito del estudio fue probar la efectividad entre tratamientos, (el método tradicional y el método digital), segundo las fuentes de invalidez externa, no son cubiertas por completo para ninguno de estos diseños y por último, la asignación al azar de los participantes a los grupos fue con el fin de homogenizarlos y en ningún momento se pretendió buscar la representatividad de población alguna. Cuando se quiere encontrar representatividad de la población estamos pisando en otro terreno.

Metodología por Encuesta

La metodología de encuestas se caracteriza por obtener muestras a partir de una población, bien definida, con el fin de extender los resultados encontrados en la muestra a la población. Dicha metodología define una serie de pasos a seguir: a) definir

los objetivos del estudio, en términos de las variables pertinentes y la población, b) implementar los instrumentos de medida, que pueden ser encuestas, inventarios, escalas... diseñados exprofeso para el caso, c) seleccionar el tamaño de la muestra, a partir de algoritmos específicos, d) seleccionar el tipo de muestreo más apropiado y que depende de las características estructurales de la población (muestreo aleatorio simple, por cuotas, estratificado...) e) recolección de la información y construcción de la base de datos, f) análisis de la información y g) presentación del informe. Esta metodología es empleada, en esencia, por investigadores de mercado, instituciones gubernamentales o políticas y en menor medida por sociólogos y psicólogos.

Esta metodología tiene como propósito central generalizar los hallazgos de la muestra a la población, por lo que no basta que los cuestionarios o pruebas estén bien desarrollados, la manera en que se elige la muestra es crucial: El cálculo del tamaño de la muestra y la elección al azar es trascendente.

Sea el caso por ejemplo, de un PO que intenta averiguar los niveles de estrés de trabajadores de la empresa QW, con el fin de implementar programas que ayuden a reducir los mismos y así incrementar la seguridad y productividad. La empresa está compuesta por ayudantes generales, obreros, personal de oficina dividido en secretarías y archivistas, trabajadores administrativos, jefes de oficina y departamento, que son en total 500 individuos. Con esta información el PO sabe que su universo o población está acotado por los 500 trabajadores. Con el fin de ahorrar recursos y tiempo el PO decide extraer una muestra de esta población para realizar su estudio de estrés, tomando en cuenta que por lo menos existían dos o tres estratos. Después de revisar la literatura y de acuerdo con las características de su población, el PO eligió el cálculo muestral a partir del algoritmo apropiado para muestras pequeñas (menor a 10 000 casos) y que permite evaluar y comparar diferentes estratos que son propios de la población (trabajadores eventuales, de planta y administrativos), ver figura 3.1.

Figura 3.1 Fórmula para calcular el tamaño de la muestra.

$$n = \frac{\frac{Z^2 q}{E^2 p}}{1 + \frac{1}{N} \left[\frac{Z^2 q}{E^2 p} - 1 \right]}$$

Donde:

Z = es el nivel de confianza y que de manera general es elegido el 95% o 99%, de seleccionar el primero quiere decir que estamos dispuestos a aceptar un 5% de error y 95% de confianza en los resultados que encontremos son representativos de la población.

p y q = a la variabilidad de el evento estudiado, cuando no se conoce ésta, que es lo más común, es asignada igual probabilidad de ocurrencia (0.5 y 0.5).

E = es la precisión con la que se generalizaran los resultados, a mayor precisión mayor es el tamaño de la muestra y visceversa, a menor precisión menor tamaño de muestra.

N = tamaño de la población.

n = tamaño de la muestra.

Donde Z se calcula a partir del nivel de confianza dado por la distribución normal.

Los valores de p y q serán de .5, debido a que no hay estudios previos que muestren que proporción de trabajadores se estresa más o menos. En cualquier caso habría que realizar una investigación bibliográfica que pudiera dar una idea para modificar estos valores, por ejemplo suponiendo que la literatura marcara que los oficinistas son más propensos al estrés que los obreros en uno 60% entonces las proporciones variarían para p y q .40 y .60.

El grado de precisión E para generalizar la muestra no es complemento de Z, de hecho E puede tomar diferentes valores por ejemplo .05, .03, .1..., al margen de Z. Entre mayor es el nivel de generalización requerido la muestra es mayor. Para nuestro caso es de .05, moderado.

Substituyendo los valores en la formula

$$n = \frac{\frac{1.96^2 \cdot .5}{.05^2 \cdot .5}}{1 + \frac{1}{500} \left[\frac{1.96^2 \cdot .5}{.05^2 \cdot .5} - 1 \right]}$$

Los cálculos muestran que n=75.68064, redondeado n= 76.

Los cálculos numéricos, y aceptando un nivel de error bajo, mostraron que 76 individuos son un tamaño muestral apropiado. El siguiente paso consiste en conocer cuántos individuos por estrato son pertinentes. El PO sabe que de los 500 empleados 430 son trabajadores y 70 oficinistas, con lo que 430 es igual al 86% de la muestra y 14% de oficinistas. Así, el 86% de 76 es igual a 65 trabajadores y 14% corresponde a 11 oficinistas.

Con el tamaño de la muestra y estratos completados el siguiente paso es elegir al azar a los participantes, lo que es posible a partir de un procedimiento simple; como partir de un listado y numerar a los trabajadores y en seguida seleccionarlos a partir de una tabla de números aleatorios o aplicación para computadora o internet (por ejemplo, <http://www.generarnumerosaleatorios.com>) donde se ofrecen programas automatizados, (para más detalle acerca de procedimientos para calcular el tamaño de la muestra ver, Lohr, 2000; Scheaffer, Mendenhall III y Lyman, 2007; Rojas, 2013). Una vez identificados los participantes se les aplica la prueba de estrés laboral y cédula de demográficos.

El PO analiza los datos con el fin de dar respuesta a sus preguntas y diseña tablas de frecuencias y porcentajes, gráficos de barras y pastel para mostrar los resultados. De la misma manera desarrolla pruebas estadísticas como "t" de Student y análisis de varianza para contrastar las muestras por actividad, edad, género... una vez hechos los análisis el PO puede tomar decisiones para implementar programas contra el estrés.

En cualquier caso, el uso de pruebas paramétricas señalan entre los principales requerimientos el muestreo al azar, independencia entre observaciones, igualdad de varianzas, la continuidad de la variable bajo estudio... y nivel de medición intervalar, esto último refiere al instrumento de medida en particular, lo que conlleva a otra área de la metodología.

Es conveniente tener presente que la aleatorización tiene funciones diferentes

en el contexto de la metodología: en los diseños de grupo cumple con el fin de homogeneizar los grupos y garantizar la validez interna y externa de los mismos, mientras que en la metodología por encuestas la muestra es extraída al azar para asegurar la representatividad de la población y poder generalizar los hallazgos muestrales a la población, con cierto grado de certidumbre.

Instrumentos de medida

En psicología, los instrumentos de medida son variados y dependen del contexto de trabajo. Por ejemplo, en el laboratorio son comunes los contadores de respuestas en función del tiempo, en el campo aplicado o de investigación existen diferentes opciones como es el caso de la metodología observacional que implica la construcción de catálogos conductuales, para a la postre, desarrollar sistemas de registro basados en frecuencias y tiempo, otro caso es la clínica donde los instrumentos son diversos y van de la entrevista, los auto registros, escalas o instrumentos de papel y lápiz. En cualquier caso, un mínimo de confiabilidad es requerida para todos estos instrumentos de medida y en particular, en otros casos, se exige la validez del mismo.

En psicología los instrumentos de medida tienen que estar sustentados por una teoría de la medida, así, la medición consiste en "... reglas para asignar símbolos a los objetos para (1) representar cantidades de atributos numéricos (escalamiento) o (2) definir si los objetos caen en la misma categoría o diferente con respecto a un atributo dado (clasificación). Ambos, escalamiento y clasificación incluyen la formulación y evaluación de reglas." (Nunnally & Berstein, 1994, Pág. 29).

En la actualidad existen dos tendencias generales en la construcción de instrumentos de evaluación, la primera es conocida como teoría clásica de los test o representacional y la segunda como la teoría de respuesta al ítem o de modelo de escalamiento. Desde el punto de vista de la teoría clásica, medir es asignar números a las observaciones de acuerdo a reglas pre establecidas, estas reglas son las escalas de medida, nominal, ordinal, intervalo y razón. El problema de esta postura radica en que el test bajo estudio tiene que demostrar, antes cualquier otra cosa, que cumple los requerimientos de la escala de medida. Así, el cero al que hacen alusión las escalas de intervalo o razón tiene que ser interpretado, y un cero de estrés qué significa, ¿ausencia de estrés o el nivel más bajo de estrés o un estrés manejable?, que si lo analizamos puede incluso ir en contra de la teoría. El otro problema está en la igualdad de los intervalos, cuando observamos un reloj o una cinta métrica es fácil y está a la vista que los intervalos son iguales, sin embargo con las pruebas o test no es posible tener esta obviedad, cuando esto ocurre (que es lo más común) se recurre a evaluar la normalidad de los datos. Si la normalidad es obtenida entonces se asume que los intervalos son iguales, de lo contrario siempre queda la duda de si la variable bajo estudio se distribuye de manera normal o tiene algún otro tipo de distribución y por eso no observamos normalidad y poco o nada podemos asumir respecto de la igualdad de los intervalos. Lo paradójico de esta situación es que en la actualidad existe una gran cantidad de instrumentos de medida donde se reportan alfas de cronbach, análisis factorial exploratorio, y análisis factoriales confirmatorios, que dicho sea de paso son análisis multivariados que asumen requerimientos de medida elevados y en ningún caso hay preocupación alguna por los intervalos iguales de los test. (para una crítica

más detallada ver Nunnally & Berstein, et. al.)

En el caso de la segunda Nunnally & Berstein, (et. al.) afirman que...

Un mucho mejor criterio es la medida en que los resultados del uso de la escala se ajustan a un modelo de escalamiento. Todo uso de la medición se basa, esencialmente, en la convención, y se avanza cuando se acuerdan mejores convenciones. En general, cuanto más bien elaborada se enuncia cuantitativamente una hipótesis, más importantes son los problemas de escalamiento formal. (Nunnally & Berstein, et. al., pag. 34).

Bajo esta postura no es necesario apegarse a la convención de escalas propuesta por Stevens, (1946). Según sea el caso, se construyen las escalas siguiendo una lógica teórica y son sometidas a pruebas (usando métodos logarítmicos, Log, Log lineal...) y en última instancia el criterio de validez está dado por su utilidad práctica, o validez empírica para probar la teoría "Independientemente de si un conjunto de ítems se ha convertido en una escala mediante procedimientos clásicos o modernos, el valor final de esa escala o cualquier otra medida se define por sus relaciones con otras variables." (Nunnally & Berstein, et.al., pag.)

No es propósito de este trabajo hacer una descripción detallada de estas posturas teóricas de la medida, para una revisión más completa sugerimos consultar Nunnally & Berstein, (et.al.), Muñiz, (2018).

Al igual que en otras disciplinas puede existir más de un instrumento, con propiedades diferentes, para medir un fenómeno, por ejemplo en la física es posible medir la longitud a partir del sistema métrico decimal o el sistema métrico inglés, lo mismo que la temperatura que puede ser medida en grados centígrados o Fahrenheit. En psicología por ejemplo es posible medir el conflicto marital a partir de entrevistas, observación conductual directa, o escalas de conflicto marital. Por otra parte, hay instrumentos que son más isomórficos que otros, por ejemplo podemos medir la longitud a partir de los pasos que damos de un lado a otro o cinta métrica, en este caso decimos que la cinta guarda mayor isomorfismo con el objeto medido que los pasos, es lógico que los pasos pueden variar en longitud dependiendo de la estatura de quien los ejecute, mientras que la regla o un flexómetro no, en todo caso el flexómetro tiene mayor isomorfismo que los pasos dados. El isomorfismo entonces se refiere a qué también el instrumento elegido representa mejor al fenómeno que queremos medir, de ahí que una entrevista puede detectar conflicto marital, mientras que la observación directa especifica el tipo de interacciones maritales conflictivas y una escala de medida ofrece algún grado de conflicto marital, además de las posibles áreas conflictivas. Así, la elección del instrumento y su adecuación isomórfica tiene que ver con el propósito de quien lo usa, ya sea con fines de diagnóstico, investigación o detección.

Cualquiera que sea el método que se elija para construir un test, al final lo importante es que de cuenta del propósito para el que fue creado, por lo que un instrumento tiene que ser confiable y válido.

Así, la confiabilidad o fiabilidad se refiere a que el instrumento de medida tiene que ser consistente, es decir, tener un mínimo de error. La consistencia puede ser evaluada a lo largo del tiempo o a lo interno. Cuando el instrumento es evaluado a lo largo del tiempo es aplicado en dos ocasiones diferentes -entre la primera y segunda aplicación tiene que haber al menos 30 días de diferencia- al mismo grupo de individuos. Los coeficientes de correlación son, en este caso, los indicadores de fiabilidad de la prueba.

Para la consistencia interna de la prueba o test es posible utilizar diferentes

índices como el alfa de Cronbach, Coeficientes de Kuder-Richardson, Rulon... que requieren una sola aplicación.

Además, la confiabilidad es esencial para que a la postre se dé la validez, pueden existir instrumentos que sean confiables pero no válidos, pero no puede haber validez sin confiabilidad.

Por otra parte, la validez puede expresarse de tres diferentes maneras, validez predictiva, validez de contenido, y validez de constructo: La validez predictiva se refiere a la efectividad del instrumento para predecir un resultado futuro, por ejemplo, cuando en las escuelas se aplica, al inicio, un instrumento para conocer que estudiantes es más probable que concluyan sus estudios, o aquellos que los abandonen, al final los resultados de esta medida son comparados con la conclusión de los estudios o la deserción. De hecho los instrumentos de orientación vocacional pertenecen a esta categoría.

La validez de contenido afirma que, de todo el posible universo de ítems se seleccionen aquellos que mejor representen al constructo, en este caso es común que se recurra a expertos, en el tema, que juzguen si los ítems seleccionados cubren el espectro del constructo en estudio.

La validez de constructo hace referencia a que el instrumento sea útil para corroborar las hipótesis derivadas del constructo – verificar la teoría-, es decir, que el instrumento y la teoría son validados a la vez (Nunnally & Berstein, et. al.). Lo anterior implica, investigación empírica que dé cuenta de la veracidad de la teoría.

Si bien es conveniente que el psicólogo organizacional tenga conocimientos acerca de la teoría de la medida, también lo es, que conozca estadística y el manejo de programas estadísticos.

Con el fin de ilustrar lo anterior, el PO está interesado en elegir un instrumento para medir el estrés laboral. El primer paso consiste en una búsqueda bibliográfica de instrumentos que miden el estrés laboral. En esta revisión el PO encuentra dos instrumentos de características similares, para decidir qué instrumento es el más idóneo se basa en dos criterios generales, la fiabilidad y la factorización propuesta por los autores. La tabla número 3.2 muestra algunos de los datos más importantes que el PO toma en cuenta para decidir qué instrumento es el más idóneo para su investigación.

	Fiabilidad del test	KMO	Factorización	Rotación	Número de Factores
Test estrés 1	.83	.81	Componentes principales	Oblimin	dos
Test estrés 2	.80	.81	Máxima verosimilitud	Varimax	dos

Tabla 3.2 Indicadores de fiabilidad y factorización

A partir de la revisión de los reportes escritos e índices de confiabilidad, KMO, factorización y rotación el PO decide elegir el test de estrés número 2. Al parecer la decisión no es muy afortunada, la lógica indica que tiene mejores niveles el primer test de estrés. Sin embargo, al revisar el PO las correlaciones ítem - total para la prueba 1 encuentra que algunos de los ítems presentan puntuaciones por abajo de .20 y otros de .90, ¿pero ésto que indica? las pautas recomendadas para las correlaciones ítem total proponen .40, con lo que aquellos ítems menores a este valor tienen que ser eliminados, lo mismo que los superiores a .90, que muestran alta redundancia. Ambos criterios son estándares generales que se siguen en la actualidad (Muñiz, et.al; Lloret, Ferreres, Hernández & Tomás, 2014), en el primer caso quiere decir que valores inferiores a .40 aportan niveles bajos de varianza explicada al test, en otras palabras,

estos ítems contribuye poco a identificar o representar el estrés -son ítems espurios-, y en el segundo la redundancia es tanta que, en un sentido estricto, es posible afirmar que hay ítems que son iguales aunque con diferente redacción, de hecho, todos los ítems contienen algún grado de redundancia -expresado en términos de correlación- por ejemplo, un ítem puede preguntar acerca de “qué tanto estrés le causa interactuar con personas del sexo opuesto” y otro ítem pregunta “qué tanto estrés le causa interactuar con mujeres”, en este caso, si quien contesta el segundo ítem es un hombre para él las mujeres son el sexo opuesto, por lo que en esencia se está preguntando lo mismo, así se obtiene una correlación muy alta o casi perfecta. De hecho, cuando se factoriza sin haber prescindido de los ítems ficticios pueden obtenerse más factores y a la vez poca consistencia dentro de los mismos, pero cuándo en la base de datos es depurada los factores son más claros y consistentes.

Otro elemento en el que pone atención el PO es la cantidad de sujetos y las opciones de respuesta de la escala. Para el primer caso observa que se reportan 180 participantes y en el segundo 215. Los estándares actuales marcan que como mínimo son necesarios 200, tanto para estimar los niveles de fiabilidad como para obtener el análisis factorial. Respecto de las opciones de respuesta, el PO observa que la primera escala cuenta con 3 opciones de respuesta mientras que la segunda con 6 (0 a 5). Diferentes autores reportan que una escala por ejemplo tipo likert debe de contener al menos 5 opciones de respuesta, lo que se justifica en términos de suponer que la variable es continua y al agregar opciones de respuesta se gana en variabilidad. (Pérez y Medrano, 2010; Lloret, Ferreres, Hernández & Tomás, et.al). En realidad, variables como la longitud son continuas pero la manera en que se implementa la medida no alcanza a cubrir la totalidad de la continuidad, así alguien puede medir de estatura 1.70 y dependiendo del flexómetro puede ser 1.702356189..., pero esta precisión depende del instrumento, en principio, y de la precisión con que se quiera medir. (Pérez y Medrano, et.al.).

Una vez refinada la base de datos, el siguiente paso consiste en decidir el tipo de análisis, Análisis Factorial Exploratorio (AFE) o Análisis Factorial Confirmatorio (AFC). Es común, cuando la escala es nueva, usar el AFE, en este caso, el primer instrumento fue factorizado a partir del Análisis de Componentes Principales (ACP) y el segundo mediante Máxima Verosimilitud (MV). El primer método fue diseñado para tomar en cuenta la varianza total mientras que el segundo sólo toma en cuenta la varianza compartida por los ítems (comunalidades) y deja de lado la varianza no compartida. Bajo ciertas circunstancias pueden darse resultados muy semejantes entre las dos maneras de factorizar, sin embargo hay una distinción conceptual muy importante,

“ Las diferencias entre la primera opción (AFE) y la segunda (ACP), son evidentes. Las variables observadas (ítems) son las variables independientes en el ACP, pero las variables dependientes en el AFE. A nivel conceptual (y formal) la diferencia es enorme. En cambio, diferentes razones han llevado durante décadas a una gran confusión entre ambas, de forma que sistemáticamente se ha aplicado una de esas técnicas, el ACP, para conseguir el objetivo que corresponde a la otra, al AFE” (Lloret, et.al., pág. 1153).

Baste un ejemplo, en una tienda de conveniencia un investigador de mercado está interesado en conocer cuáles son los indicadores o variables que pueden ayudar a entender la preferencia por comprar un jabón de baño en particular. Con tal fin, desarrolla una serie de ítems donde pregunta a los participantes, seleccionados, acerca de su decisión. Las preguntas se relacionan con la fragancia, el precio, presentación, cantidad de producto (en gramos), qué beneficios ofrece, etcétera. En seguida se

procesa la información para obtener los componentes (ACP) que van a probabilizar la elección del tipo de jabón que compra la gente (por ejemplo, primer componente propiedades benéficas del jabón y segundo componente precio - disponibilidad). Es decir, los componentes sirven como variables predictoras (independientes) de la compra del tipo de jabón.

Para el caso del AFE, la técnica (Máxima Verosimilitud) toma en cuenta sólo la varianza compartida entre los items, lo que hace de entrada, que los items que se agrupen en factores que describen, de la mejor manera, al constructo de interés, el estrés, la variable dependiente. En este caso la técnica elegida exige normalidad, independencia en las observaciones y que no exista colinealidad. Cuando los supuestos no son satisfechos hay otras aproximaciones para factorizar (Hair, Anderson, Tatham & Black, 1999; Lloret, et.al., 2010).

De esta manera, el PO toma la decisión basado en los estándares recomendados, con lo cual gana en certeza de que los resultados que encuentra en su aplicación tienen un alto grado de confianza. Así, sus conclusiones y afirmaciones están bien fundamentadas. No hacer caso de los estándares conlleva a resultados inciertos.

Comentarios

El propósito del presente escrito fue hacer una breve exposición sobre algunos de los problemas metodológicos que más llaman la atención a los estudiantes de posgrado. Creemos que es necesario, si no indispensable, que los psicólogos tengan una identidad propia en los diferentes ámbitos de trabajo, la razón básica estriba en que en muchas ocasiones los psicólogos tienden a mimetizarse en los ambientes laborales: cuando el psicólogo trabaja en hospitales adopta un lenguaje similar a los practicantes de la medicina, lo mismo ocurre en las empresas, en este caso el psicólogo habla de la productividad, las ganancias, si tal o cual cosa le es útil a la empresa, su pensamiento y lenguaje tiene un sentido económico administrativo empresarial. Por otra parte, cuando preguntamos a otros profesionales a qué se dedica el psicólogo sus respuestas son ambiguas, poco claras, y desinformadas, respuestas del estilo de “se dedican a lo mental”, “a entrenar ratas o pichones”, “al sentido humano de la vida”, son expresiones que muestran el poco conocimiento del quehacer de el psicólogo, esto no es otra cosa que falta de identidad. Lo que tiene que darle al psicólogo identidad propia es su educación, una formación teórica, metodológica y práctica. El psicólogo tiene que asumir que su materia de estudio es el comportamiento humano individual y para estudiarlo cuenta con estrategias metodológicas diversas, lo presentado en éste escrito es una muestra, desde la visión metodológica.

La recomendación es conocer la metodología que se elige para responder sus preguntas o necesidades prácticas. La elección de la metodología de grupos, por encuestas y psicometría está fundamentada por razones de carácter práctico, esta metodología permite hacer investigación en ámbitos como, la educación, salud y empresas, lo mismo puede contribuir a resolver problemas concretos que ocurren en las organizaciones y en el mejor de los casos hacer investigación aplicada.

En lo referente a la metodología de grupos y por encuestas, la pretensión fue marcar las diferencias de propósitos, a partir del azar, definido como que todos los elementos de una población tengan la misma probabilidad de ser elegidos. En la metodología de grupos el azar tiene como propósito homogeneizar los mismos y a partir

de esto es que son declarados como diseños pre experimentales o experimentales. Así, cualquier manipulación sobre la variable de interés o dependiente permite, con mayor confianza, afirmar que es debida al tratamiento y no producto de la casualidad.

"Campbell (1957) distingue dos tipos principales de validez en los diseños experimentales. ... por tanto, en la validez interna la efectiva acción de los tratamientos experimentales que debiera reflejarse en la diversificación de los resultados. El propósito principal de todo investigador es, sin duda alguna, conseguir dicha validez, ya que gracias a ellas puede lograr una comprensión empírica de sus presupuestos hipotéticos. Cuando, por otra parte, habla de la validez externa de un diseño, Campbell se refiere a la representatividad o generalidad de los resultados de un experimento. Dicha validez se relaciona, por consiguiente, a las clases de poblaciones a las que pueden generalizarse los resultados. Y estas poblaciones no deben limitarse exclusivamente a las de los sujetos, sino que incluyen las relativas a las de los tratamientos e incluso a las de las <pautas o patrones estimulativos" (Arnau, 1981, Pág. 9)

En tanto, en la metodología de encuestas el azar tiene como propósito la representatividad de la población, así, los hallazgos muestrales pueden ser generalizados a la población.

"El muestreo probabilístico es la base fundamental de cualquier investigación mediante encuesta... la clave para conseguir este objetivo es un principio fundamental denominado "igual probabilidad de selección", que afirma que si cualquier miembro de una población tiene la misma probabilidad de ser seleccionado en una muestra, entonces esa muestra será representativa de la población. Es así de sencillo." (Scheaffer, Mendenhall & Lyman, 2007, Pág. 16).

En cuanto a las propiedades psicométricas de los instrumentos, es conveniente que los interesados en aplicar test tengan en cuenta la importancia de conocer los aspectos teóricos, estadísticos e interpretativos de la confiabilidad y validez, tanto para elegir o diseñar una escala. La confiabilidad, determinación de los factores (AFE) y verificación de la estructura de varianza (AFC) son elementos necesarios para iniciar la verificación de las hipótesis derivadas del constructo, es decir, probar la teoría.

Referencias

- Arnau, G.J. (1981). Diseños experimentales en psicología y educación. Métodos y técnicas de investigación psicológica. Volumen 1. Ed. Trillas. México.
- Campbell, D.T Y Stanley, J.C. (1995). Diseños experimentales y cuasiexperimentales en la investigación social. Ed. Amorrortu. Buenos Aires.
- Gómez-Benito, J. M., Hidalgo, M.D. y Guilera, G. (2010). El sesgo de los instrumentos de medición. Tests justos. Papeles del Psicólogo. Vol. 31(1), pp. 75-84.
- Hair, Jr.J.F., Rolph E. Anderson, E.A., Tatham, R.L. y Black, W.C. (1999). Análisis multivariante. Prentice Hall. México.
- Lloret-Segura, S, Ferreres-Traver, A., Hernández-Baeza, A. y Tomás-Marco, I. (2004). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. Anales de Psicología, 2014, vol. 30, no 3 (octubre), 1151-1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>.
- Lohr, S.L. (2000). Muestreo diseño y análisis. Ed. Thomson. México.
- Muñiz, F.J. (2010). Las teorías de los tests: Teoría clásica y teoría de respuesta a los ítems. Papeles del Psicólogo. Vol. 31(1), pp. 57-66.
- Muñiz, F.J. (2018). Introducción a la psicometría. Teoría clásica y TRI. Ed. Pirámide. España.
- Nunnally, J.C. y Bernstein, I.J. (1995). Teoría psicométrica. McGraw Hill. México.
- Pérez, E.R. y Medrano, L. (2010). Análisis factorial exploratorio: bases conceptuales y metodológicas. Revista Argentina de Ciencias del Comportamiento, Vol 2 N 1, 58-66.
- Rojas, S.R. (2013). Guía para realizar investigaciones sociales. Ed. Plaza y Valdés. México.
- Scheaffer, R.L, Mendenhall III, W. y Lyman, O.R. (2007). Elementos de muestreo. Ed. Thomson. México.
- Stevens, S.S. (1946). Sobre la Teoría de las Escalas de Medición. Science. Vol. 103, No. 2684.



Parte IEl Rol del Psicólogo
Organizacional**1** **2** **3** **4** **5****Parte II**Gestión Organizacional
Basada en Evidencias**6** **7** **8** **9** **10****Parte III**Psicología Aplicada al
Emprendimiento y a la
Gestión Organizacional**11** **12** **13** **14**