

EL USO DE TESTS PSICOLÓGICOS EN RECURSOS HUMANOS

Abelardo Pithod

Profesor Titular de Comportamiento Organizacional
de la Facultad de Ciencias Económicas
de la Universidad Nacional de Cuyo

El sociólogo ruso Pitirim Sorokin (1964), que realizó lo más importante de su obra en Harvard después de la revolución bolchevique de 1917, acuñó en la época de pleno auge de los tests y las mediciones en ciencias humanas, los calificativos peyorativos de *quantofrenia* y *testocracia*. Libro apasionado respecto del cual, a tantos años, podemos realizar una ponderación ni tan negativa como la de Sorokin ni tan optimista como la de la psicometría. Los tests se usan en Recursos Humanos -y en tantas otras disciplinas psico-sociales- y se seguirán usando. Lo malo no está en el uso sino en el abuso, en la aplicación indiscriminada y literal y no en la interpretación prudente de sus resultados.

Una interpretación prudente, o *sabia*, significa:

1. Saber qué miden los tests.
2. Conocer teoría de la medida, lo que significa tener en claro la confiabilidad y validez del instrumento que se usa.
3. Tener en claro la teoría psicológica (o sociológica) subyacente al test.

4. Saber que el error de medida de un test se refiere a *grupos* de sujetos, y que su valor predictivo (porque de predecir se trata) es mucho más azaroso cuando se refiere a sujetos individuales.

En nuestro análisis no vamos a seguir este orden temático, sino más bien un orden cronológico empírico, siguiendo los pasos del psicómetra.

Qué se espera cuando se aplica uno o varios tests (batería)

Cuando una empresa nos pide que le seleccionemos a X número de obreros con condiciones para realizar la tarea Z , podemos proceder de diferentes maneras. Podemos poner a realizar la tarea durante cierto tiempo a cada uno de los candidatos, lo que tiene el inconveniente de los costos y de que puede haber candidatos que **aún** no están prácticos en esa tarea pero que tienen potencialidades sobresalientes. Imaginemos que se trate de una tarea similar a la de tornería. Tengo 30 candidatos y solo necesito 7 sujetos. Obviamente el empresario quiere que le elija los mejores 7 entre los treinta. Como los psicólogos han estudiado las principales **aptitudes** para llegar a hacer bien o muy bien las tareas vinculadas a la tornería y han construido instrumentos de medida, someto a testificación a los 30 sujetos.

Los tests de aptitudes son los que logran un error de medida más bajo, comparados con los tests de personalidad. Siempre lo tienen, por supuesto, como sucede aún en ciencias exactas (física, geología, etc.). Ese error me dice que lograré con seguridad escoger a los 7 mejores, sin muchas dudas, pero me dice también que habrá dos (2) entre los que elija que no serán del grupo de los mejores. Que habrá en los 23 descartados dos potencialmente mejores que yo no detecté. El drama es que yo no sé cuáles son los 2 entre los mejores que no son realmente mejores. El empresario quedará contento si yo le digo: "Mire, éstos son los 7 que resultaron mejores", pero se pondrá serio si yo agrego "pero hay dos (2) con los que no estoy seguro". El contratante me dirá "¿Cuáles son esos 2?", y yo le tendré que decir: "no lo sé, pueden ser cualquiera de los 7". Lamentablemente, para desprestigio de la psicometría, los psicólogos o administradores de tests, no lo dicen, o no lo saben (esta es la hipótesis más plausible), y emiten juicios dogmáticos porque en realidad no conocen los límites de su técnica.

Hay criterios prácticos que se deben conocer para usar tests. Por ejemplo: Los derivados de la fiabilidad o confiabilidad. Uno puede tomar un test de lo que se llama "inteligencia general" o factor "G". Algunos sujetos no rinden lo que pudieran porque son medidas puntuales, y el testificando puede hallarse nervioso, o simplemente en un mal día. Su rendimiento será menor que el alcanzable. Pero la inversa no es verdadera: Si alguien descolgó en una prueba (una buena prueba) de factor "G", ese tal seguro que tiene alto ese factor (cuya naturaleza intrínseca o psicológica deberíamos conocer; de eso hablaremos enseguida). ¿Cómo es esto? En realidad es sencillo: si alguien salta una valla de dos metros, aunque sea una sola vez en su vida, ese tal es capaz de saltar dos metros. Pero otro que no lo logra no quiere decir que no pueda lograrlo en el futuro. Los psicómetros suelen olvidar un principio tan evidente. Y cuando pasan sus informes dicen, taxativamente, fulano es mediano, mengano menos que mediano y perengano es excepcional. De este último podemos estar prácticamente seguros. De los otros dos, mucho menos seguros. *Pero con una salvedad muy grave que no se toma en cuenta, y sobre la que deberemos volver.*

De estas consideraciones podemos sacar un criterio general: Cuidado con los resultados de los tests cuando las circunstancias inducen a comportamientos anómalos, por ejemplo en selección de personal, gente que se traba, que es lenta, pero que no carece de talento, etc. Igual digamos de los cuestionarios de personalidad cuando el individuo está fuertemente tentado a dar "una buena impresión". Por cierto que los cuestionarios de personalidad tienen preguntas de control de la sinceridad con que el sujeto responde, pero el saber que el sujeto no fue del todo transparente sólo me informa que no debo considerar válido ese test, y proceder a descartarlo, y usar otro, pero estaré nuevamente ante el problema de la sinceridad, conciente o inconciente. En la práctica, esto significa más tiempo, es decir, más gasto.

Lo que miden los tests

Y aquí nos sale al paso el gran tema, no cerrado todavía, de lo que miden realmente los tests. Antes de entrar en él digamos que, correlativamente, el psicómetro debe saber aproximadamente cuáles **aptitudes** son las más necesarias para desempeñar una determinada tarea. Porque el conocimiento de los factores principales requeridos para realizar cualquier tarea es algo tan necesario de saber

psicómetra debe saber aproximadamente cuáles **aptitudes** son las más necesarias para desempeñar una determinada tarea. Porque el conocimiento de los factores principales requeridos para realizar cualquier tarea es algo tan necesario de saber como lo que miden los tests. En efecto, las tareas son siempre actividades complejas. Nunca sabemos del todo bien qué **constelación de aptitudes** es la ideal para cada una o si diversas constelaciones pueden hacer aptos a distintos sujetos, aunque con estilos distintos. Un ejemplo: Hay sujetos que son muy rápidos en actividades psicomotrices -por ejemplo, escribir a máquina y ahora teclear un procesador de textos-, pero puede que uno no tan rápido no cometa tantos errores, con lo que la simple velocidad queda compensada por la exactitud. Otro: puede que un sujeto sea desparejo en sus performances, que haya días que cometa muchos errores y otros días pocos. Hay sujetos que tienen poco rendimiento en ejercicios de memoria, pero que lo compensan con creatividad o capacidad de improvisación (que es, sencillamente, rapidez de adaptación). Habrá que ponderar qué nos conviene más para determinado puesto.

Observe el lector que aún no hacemos referencia a las condiciones de **carácter**, también llamadas de **personalidad**. Hemos aludido solamente a las aptitudinales.

El tiempo también afecta la confiabilidad de los tests. Hace años que se usa el Test de Matrices Progresivas de Raven para factor "G" (inteligencia general). Nuestra experiencia es que este instrumento tiene dificultades: algunos han señalado que no está tan libre de influencias culturales como creyó su autor. Pero últimamente, a nosotros, nos está pasando algo más grave. Por lo menos en sujetos universitarios el test presenta problemas. Ante todo resulta demasiado fácil para algunos sujetos, con lo que sus puntuaciones los colocan, respecto de los baremos nacionales, en posiciones muy altas. Tenemos casos de sujetos que responden en 20 minutos **todos** los ejercicios y acertadamente, con lo cual se convierten en genios psicométricos, o al menos en superdotados, y en realidad no lo son. Hemos contrastado Raven con Cattell y éste no presenta distorsiones en el polígono de frecuencias. ¿Qué significa ésto estadísticamente? Algo sencillo: el test de Raven no está discriminando bien. ¿Por qué? Porque ningún test de aptitudes, cualquiera que sean ellas, debe permitir que algún sujeto -y menos que varios- alcancen la puntuación máxima posible. Así como un test bien hecho debe tener preguntas muy fáciles al comienzo para descargar la

ansiedad de los sujetos y motivarlos a seguir, no debe dejar llegar, salvo por rarísima excepción, al máximo puntaje posible. Un test de aptitudes debe dar una distribución de frecuencias aproximadamente semejante a la curva normal de probabilidades, es decir en forma de campana. Y con el Raven nos está sucediendo que presenta una asimetría de la curva volcada hacia la derecha (es decir, es demasiado fácil, y, además, poco discriminante entre los que alcanzan puntuaciones altas, que son demasiados). ¿Cómo explicamos este fenómeno del Raven? No tenemos hipótesis ni estudios al respecto. Sólo nos animaríamos a pensar, por ahora, que puede estar habiendo influencias culturales: las computadoras y, quizá, la enseñanza de la llamada matemática moderna. Pero no lo sabemos. Sería un interesante tema de investigación. El test de Raven consiste en matrices a resolver. Algunos sujetos con estudios en matemática (¿tendrá que ver el álgebra de matrices?), resuelven algunas de las preguntas del Raven por su similitud con lo que saben de álgebra matricial. Otros ítems -he aquí un problema diferente- son resueltos intuitivamente, casi por un juicio estético. Se dice que en un alto nivel matemático no estaría ausente un cierto factor estético, pero no parece que sea éste el caso. Por lo menos no lo sabemos. Las "matrices progresivas" del Raven pueden ser captadas holísticamente, en visión intuitiva, como la del artista, y entonces no estamos frente a factor "G" (inteligencia abstracta) en estado puro, que es lo que dice medir el test. El test está por lo menos contaminado. Cattell intenta apreciar también factor "G" y no parece presentar estos inconvenientes, pero no lo podemos asegurar todavía porque nuestra experiencia con él es menor.

Hechas estas observaciones podemos pasar a ver qué tipos de tests se usan más comúnmente entre nosotros en Recursos Humanos.

Tipos de tests y su uso

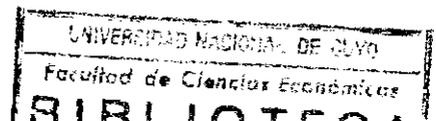
Hay abundante bibliografía sobre el tema (cf. Referencias Bibliográficas). No es necesario, tampoco, que repasemos la historia de la psicometría, por lo demás llena de enseñanzas y, también, de sorpresas. Aquella famosa frase, atribuida a uno u otro de los pioneros de la psicometría, cuando se le preguntó qué era la inteligencia y se limitó a responder: Lo que miden los tests, no vaya a creerse que era una simple "boutade". La dificultad no está definitivamente superada como etapa del estudio de

la mente. Hasta bien avanzado el análisis factorial, en realidad lo que los tests medían eran habilidades más o menos específicas para realizar ciertas tareas, desde la comprensión de textos y razonamiento verbal, hasta la aptitud psicomotriz de un sujeto, pasando por la memoria, la resolución de problemas espaciales, el descubrir la ley que preside una serie de números o de letras, y tantas otras.

No es necesario volver a la polémica de si realmente existe ese famoso factor "G" o de inteligencia general que postuló la escuela inglesa de psicometría, desde los tiempos de Spearman (1904; 1927; trad. cast. 1955) y que resistieron los investigadores americanos, Thurstone (1938) y otros muchos. En concreto se siguen usando tests de factor "G" y tests más específicos (inteligencia perceptual-intuitiva, psicomotriz, etc.). El propio Cattell, hombre de la línea norteamericana, ofrece, como vimos, algunos tests de factor "G", aunque no un único test, como el de Raven, sino cuatro (4). El test de Raven tiene la ventaja de no provocar prisa ni ansiedad porque no es de tiempo estricto. Cattell es un test de tiempo. Conocemos la objeción de los que están en contra de tests con límite de tiempo, sobre todo cuando las limitaciones son muy estrictas (los 4 tests de Cattell exigen un máximo de 3 minutos para el test 1, cuatro para el 2, 3 para el 3 y 2,30 para el 4, en total 12 minutos y 30 segundos). Pero puede compensarse esta desventaja observando el comportamiento de los sujetos durante la aplicación y controlando las muestras de nerviosismo, inhibición, perplejidad, lentitud en los más reflexivos y menos intuitivos, exacerbación de las dudas en los obsesivos, en una palabra teniendo bien en cuenta las circunstancias de aplicación. Si se trata de una selección, son de esperar distorsiones; se debe, pues, determinar el límite de tolerancia de estas distorsiones en cada situación, según las consecuencias y demás variables que el aplicador de tests debe tener en claro. Muchas más veces de lo que se quisiera, los psicómetros o administradores suelen mecanizar su tarea, por motivos pragmáticos, y eliminar *a priori* las diferencias individuales frente a estados de estrés. Esto es muy poco "psicológico". Obviamente que es más económico hacer caso omiso de las limitaciones de los tests en cada situación, pero el error de medida puede aumentar alarmantemente y quedar frustrados no solo los sujetos sino quien pide la testificación, en nuestro caso los empleadores.

Tests para predicción académica y tests para rendimiento laboral

Hay que saber que los tests que se comercializan más comúnmente entre nosotros han surgido de la necesidad de predicción del rendimiento académico, lo cual no es por cierto lo mismo que la predicción laboral. Hay una generalizada sensación, basada en suficiente evidencia empírica, en el sentido de que el éxito académico no se corresponde exactamente con el éxito laboral. Y "ciertamente, los grados académicos no son una medida adecuada para (predecir) el éxito en la vida" (Sternberg, 1990, 287) y por ende que tests concebidos para predecir éxito académico no predicen bien la eficiencia laboral. Es totalmente lógico. Las aptitudes necesarias en uno y otro caso pueden ser disímiles o influir factores extra-aptitudinales, lo cual puede hacer que un sujeto académicamente brillante (un medalla de oro), sea en la actividad profesional menos bueno que un sujeto que compensa su menor éxito en los estudios con condiciones de otro tipo, sean de carácter, como el buen rapport social, las aptitudes para la conducción, poder de síntesis, y otros muchos factores aún no bien conocidos y que se vinculan a la productividad o éxito prácticos. Decimos "aún no bien conocidos" porque entendemos bien cómo diferentes "constelaciones" de rasgos pueden provocar buenos efectos, es decir, personalidades disímiles lograr éxito aún siendo muy distintas entre ellas. Hay sujetos a los que, siguiendo "patterns" académicos o psicométricos no podemos calificar de "talentosos" o "inteligentes", pero que están dotados de cierta astucia o agudeza práctica para el manejo de negocios o personas. Estas capacidades, llamémoslas "intuitivas", no se pueden poner en evidencia mediante tests convencionales de aptitud. Aquí vale el dicho criollo: "en la cancha se ven los pingos". De éstas y otras razones vamos a derivar nuestra propuesta básica: No separar selección de capacitación, es decir, lograr en lo posible que los exámenes de selección inserten de algún modo la observación de los sujetos "en la cancha", luego de un breve e intensivo ejercicio sobre el terreno observar cómo resuelven problemas concretos. Este mayor "gasto", en realidad es una inversión. La selección demasiado artificial y aparentemente más expeditiva puede dar origen a errores lamentables, cuyos costos serán mayores a la larga.



Tests estáticos y capacidad potencial

Cada vez se acepta menos que los tests de cociente intelectual o de habilidades más específicas sean inamovibles en sus resultados. Es decir, pueden ser modificados por el entrenamiento. De hecho ahora sabemos que la intensidad del aprendizaje a que se ven sometidos los chicos en Japón ha elevado en alrededor de tres años el CI de la población infantil respecto, por ejemplo, de la población inglesa. El CI (cociente intelectual) se obtiene dividiendo la edad mental (o psicométrica) con la edad cronológica. Así, si un niño obtiene en los tests resultados superiores a su edad cronológica, el cociente será mayor que 1, lo cual multiplicado por 100 nos da el CI. Si un niño de 5 años cronológicos obtiene puntuaciones medias de chicos de 10 años, ese niño tendrá un CI de 200, es decir, será un "genio" psicométrico (no confundir con el genio sin más). Si obtiene alrededor de 100, será un niño promedio, y si baja sensiblemente de 100 será considerado "fronterizo" y si baja aún más, "deficiente". Hoy se habla de discapacitados o "especiales".

Pero volvamos al tema de si se puede o no hacer crecer el CI mediante entrenamiento.

Tests académicos y tests laborales

Vigotsky (1978) y Feuerstein (1980) han señalado vigorosamente cómo la ejercitación o entrenamiento pueden modificar el potencial intelectual o aptitudinal de las personas. El mito del cociente intelectual (CI) estable a partir de los 14 o 15 años, medido de manera constante por los tests clásicos de CI (v.g. los tests de factor "G") no resiste, al parecer, la crítica reciente (Gardner, 1983; Sternberg, 1990, entre otros). Obviamente esta crítica choca contra intereses creados en torno a la industria de los tests. Si los tests de CI y los de habilidades específicas no se salvan de la modificabilidad por aprendizaje, la predictibilidad se torna problemática. Por cierto que son aprendizajes lentos y no invalidan de manera absoluta las medidas puntuales de los tests¹.

¹ Habilidades específicas primarias *-primary mental abilities* las llamó Thurstone (1938)-son inteligencia verbal, de tanta relevancia en el aprendizaje escolar y académico, intelligen-

Conclusiones

1. En Recursos Humanos, tanto psicólogos como administradores, seguimos trabajando con instrumentos a los que caben todas las restricciones que acabamos de señalar. Siempre que las restricciones de tiempo y costos no permitan otra cosa, debemos ser concientes de la real limitación en cuanto a confiabilidad y validez de nuestros tests como predictores de éxito laboral (o de algunos aspectos de él).

2. Es urgente actualizarnos recurriendo a los centros de investigación psicométrica y/o de Recursos Humanos de avanzada en países desarrollados. Un aspecto relevante es conocer la metodología que se está usando para captar las **potenciales** de aprendizaje de los sujetos, dato más importante que la medida puntual válida para el aquí y ahora. Conocer las potencialidades de aprendizaje es hoy más significativo que la presente situación de los sujetos. Un dato muy ilustrativo: el entrenamiento ha hecho que en Israel Reuven Feuerstein haya logrado notables resultados con niños que padecen síndrome de Dawn. Algunos están escolarizados en escuelas comunes y Feuerstein confía en que su nieto Dawn llegue hasta la Universidad. Sea como fuere de esto último (cuando se lo conté al gran especialista en el tema, descubridor del gene que produce el síndrome aludido, Jérôme Lejeune, me contestó algo escéptico: "¡tanto como la Universidad!") hay hechos importantes en materia de potenciar el desarrollo mental. Por ejemplo, en Israel entrenan a los Dawn para cuidar ancianos y parece que con éxito. En este marco quisiéramos que se interpretara la siguiente conclusión:

3. Nos parece que dada la movilidad de habilidades que está imponiendo el mercado laboral, debido a cambios culturales y tecnológicos, la selección de recursos humanos debe estar cada vez más ligada a la **capacitación**. Elegir a los que muestran más capacidad de aprendizaje, es decir más potencialidades, y no tanto a los que aquí

cia psicomotriz, factor decisivo en actividades prácticas o deportivas (Maradona y Pelé la poseyeron en grado superlativo) numérica, espacial (indispensable para arquitectos, ingenieros, geógrafos, etc.), mnemónica, y otras.

ahora (al aplicarles tests clásicos) obtengan buenos resultados. Este es un desideratum, quizá una utopía. Pero al menos sepamos que tal sería la situación ideal, la cual, en algunos casos al menos, puede hacerse factible, sobre todo si se trata de elegir mandos medios y altos.

4. Por último: Aspirar en todo lo que se pueda a que las pruebas de aptitud para determinados cargos se logre imiten situaciones parecidas a la realidad de dichos cargos. Y que si sólo podemos aplicar tests de aptitudes (aquí no nos ocupamos de los caracterológicos o de personalidad), sean evaluados prudentemente, es decir sabiendo lo que estamos haciendo y balanceando sus resultados con los restantes datos de la persona, antecedentes, recomendaciones, si se puede períodos de prueba de campo, etc.

REFERENCIAS

- CATTELL, R.B., *Test de Factor "G"*, Buenos Aires, Management Center, 1977.
- CRONBACH, L.J., *Fundamentos de la exploración psicológica*, Madrid, Biblioteca Nueva, 1963.
- FEUERSTEIN, R., *Instrumental enrichment: An intervention Program for cognitive modifiability*, Baltimore MD: University Park, 1980.
- GARDNER, H., *Frames of Mind: The theory of multiple intelligences*, N.York, Basic, 1983.
- GUILFORD, J.P., *La naturaleza de la inteligencia humana*, Buenos Aires, Paidós, 1977.
- GUILFORD, J.P., *Psychometric Methods*, N.York, McGraw-Hill, 1954.
- OLLERON, P., *Les composantes de l'intelligence d'apres les recherches factorielles*, Paris, PUF, 1957.
- SOROKIN, P., *Achaques y manías de la sociología moderna y ciencias afines*, Madrid, Aguilar, 1964.
- SPEARMAN, Ch., *Las habilidades del hombre*, Buenos Aires, Paidós, 1955.
- STERNBERG, R.J., *Metaphors of Mind. Conceptions of the Nature of Intelligence*, Cambridge University Press, 1990.
- THURSTONE, L.L., *Primary Mental Abilities*, University of Chicago Press, 1938.
- VYGOSKY, L.S., *Mind and Society: The development of higher psychological processes*, Harvard University Press, 1978.
- WESCHLER, D., *Test de Inteligencia para adultos*, Buenos Aires, Paidós, 1984.