

Autoinformes y respuestas sesgadas

C. DE LAS CUEVAS Y CATRESANA, J. L. GONZALEZ DE RIVERA Y REVUELTA

Cátedra de Psiquiatría. Facultad de Medicina. Universidad de La Laguna. Islas Canarias

RESUMEN

Los autoinformes constituyen un método adecuado y directo en la evaluación de las respuestas cognitivas y experiencia subjetiva del individuo. A pesar de que han sido puestas de manifiesto diversas fuentes de error en este tipo de instrumentos, existen pocas razones para pensar que las evaluaciones suministradas por éstos son menos válidas o fiables que las correspondientes a otras técnicas estandarizadas. En el presente trabajo procedemos a la descripción de los distintos errores y respuestas sesgadas de los autoinformes como método de evaluación psicológica así como de las distintas formas de localizar o identificar cada sesgo específico.

PALABRAS CLAVE; Autoinformes. Inventarios Autoadministrados. Respuestas Sesgadas.

El prototipo de todo inventario autoadministrado lo constituye la "Hoja de Datos Personales de Woodworth" (1918) desarrollada para su uso durante la primera guerra mundial.

Este autor introdujo una nueva modalidad de medida psicológica, el "autoinforme" (self-report), que posibilita el que cada individuo pueda "entrevistarse a sí mismo". Este inventario consistió esencialmente en un intento de estandarización de una entrevista psiquiátrica con vistas a la adaptación del procedimiento para la evaluación de grandes masas de individuos. Para la elaboración de su inventario, Woodworth recogió información relacionada con los síntomas neuróticos y preneuróticos más frecuentes en la literatura psiquiátrica de la época, así como procedente de conferencias que mantuvo con diversos psiquiatras, siendo las preguntas del inventario originalmente formuladas en base a esos síntomas. Las preguntas se referían a conductas desviadas, tales como miedos anormales o fobias, obsesiones y compulsiones, pesadillas y otros trastornos del sueño, fatiga excesiva y otros síntomas psicósomáticos, sensación de irrealidad y trastornos motores del tipo de los tics y temblores. En la selección final de los ítems el autor aplicó ciertas comprobaciones estadísticas empíricas, sin embargo, es manifiesto que el principal énfasis en la construcción y uso de su inventario estuvo situado en la validez de contenido, como lo indican el origen de dichos ítems, así como el reconocimiento de ciertas conductas como inadaptadas (Anastasi, 1982).

El inventario autoadministrado como método de evaluación

SUMMARY

Self-reports are an adequate and direct method to assess the cognitive responses and subjective experience of a subject. Self-report mode of assessment of the psychological status remains a useful form of clinical measurement. It can be seen that, while many sources of bias have been identified on self-ratings, there is a little reason to believe that self-ratings are any less valid or reliable than corresponding standardized techniques. In the present paper the biases and errors in self-ratings are described and how to locate each bias is considered.

KEY WORDS: Self-report. Errors and Biases. Psychological Assessment.

psicológico posee diversas ventajas y características que lo hacen recomendable para determinados propósitos, no obstante esta modalidad de medida asimismo posee ciertas debilidades o defectos inherentes.

La primera ventaja, y especialmente en relación con la psicopatología, es que permite el acceso a una información exclusiva de la cual no se puede disponer por otros medios de evaluación, pues al ser autoevaluado suministra información que proviene "desde el mismo que la experimenta" (Derogatis y Melisaratos, 1983), mientras que las observaciones clínicas se limitan a la "versión aparente" de la experiencia del paciente, basada en la conducta observable y en los informes verbales (Derogatis et al., 1974). En segundo lugar, el uso de autoinformes puede suponer un considerable ahorro de tiempo por parte del profesional que los emplee. La técnica puede ser integrada con facilidad en las rutinas institucionales y es tratable con métodos actuariales de puntuación e interpretación. Los datos que surgen de estas medidas pueden ser también incorporados directamente en los sistemas de decisión clínica (Fowler, 1969; Lanyon, 1972), y los inventarios autoadministrados han demostrado su sensibilidad a una amplia variedad de intervenciones terapéuticas. Además, como Skilbech et al. (1984) afirman, son difícilmente sustituibles si queremos conocer las diferentes formas de percibir e informar sobre síntomas psiquiátricos de los distintos grupos étnicos (Yamamoto et al., 1981; Platt et al., 1989) o clases sociales (Derogatis et al., 1971a, 1971b). Wilson et al. (1985) opinan que este tipo de inventarios autoadministrados de síntomas están diseñados para medir cuatro aspectos o facetas distintas: 1. Los cambios sintomáticos especialmente en respuesta a las diferentes posibilidades

de tratamiento; 2. La gravedad del trastorno; 3. Posibilitar o facilitar el diagnóstico, incluyéndose aquí la capacidad del cuestionario para diferenciar entre pacientes neuróticos y pacientes psicóticos, y a su vez la capacidad para diferenciar entre pacientes y personas sanas; y 4. Poner de manifiesto rasgos de personalidad.

Sin embargo, debemos considerar que el método autoadministrado tácitamente asume la validez de la "premisa del inventario", que implica que el individuo que responde puede y quiere describir de la forma más exacta posible sus síntomas y conducta actual. Esta es una premisa que no siempre se cumple, haciendo, por ejemplo, inoperante este tipo de instrumento en los pacientes psicóticos agudizados, en los dementes seniles incapaces de la comprensión de los ítems (Snaithe, 1981), y en aquellos contextos en los cuales el engaño o a la simulación puedan estar presentes (Lees-Halley, 1989), como en la simulación deliberada o exageración de una enfermedad o incapacidad para evitar una situación displacentera o para obtener una ganancia personal.

El inventario autoadministrado también se encuentra sometido a otras posibles distorsiones sistemáticas secundarias a las llamadas respuestas sesgadas que a continuación explicaremos con más detalle.

Es importante hacer la distinción entre el uso de evaluaciones para obtener información sobre el mundo exterior y su uso para medir creencias, sensaciones y emociones. Una distinción paralela es necesaria entre los sesgos en la creación de actitudes, juicios y percepciones y los sesgos que se producen a la hora de traducir o plasmar éstos en una respuesta. Cuando utilizamos evaluaciones para medir creencias, sensaciones o emociones per se debemos tener en mente este último tipo de sesgo. Desafortunadamente, sin embargo, en trabajos previos sobre sesgos de respuestas nunca ha quedado claro donde se localizan éstos (West, 1983), por eso en los siguientes párrafos intentaremos considerar cómo se puede localizar o identificar cada sesgo específico.

1. Efecto Halo

El efecto halo ya fue puesto de manifiesto por los psicólogos de los años 20 (Thorndike, 1920). Esta respuesta sesgada tiene lugar cuando el que responde generaliza en exceso acerca del objeto a valorar, de forma que valoraciones del mismo objeto en dimensiones diferentes están altamente correlacionadas. Los que responden parecen estar utilizando sus evaluaciones en una dimensión para determinar sus evaluaciones en otras dimensiones. Si el origen de este sesgo yace probablemente en los juicios mismos su grado o extensión sería equitativamente constante a través de diferentes formatos de respuesta. Es posible, sin embargo, que el efecto halo esté creado o fomentado por la ambigüedad asociada con posiciones específicas de las escalas de valoración. Allí donde el juicio no pueda ser utilizado para hacer la última elección en la valoración, otras indicaciones o ejemplos como evaluaciones pasadas pueden llegar a ser prominentes. Esto puede contar para las altas intercorrelaciones entre las evaluaciones en diferentes dimensiones.

El efecto halo puede ser cuantificado de diferentes formas (Saal et al., 1980). Una de ellas puede ser el comprobar el grado de correlación entre las diferentes dimensiones. Altas intercorrelaciones sugieren el efecto halo. También se puede utilizar la matriz de correlación como base para un análisis factorial o un análisis de componentes principales, cuantos menos factores o componentes significativos mayor será la probabili-

dad de la existencia de un efecto halo. Además podemos valorar la interacción entre evaluaciones en un análisis de varianza (ANOVA). Una interacción significativa sugiere que diferentes evaluadores de forma consistente evalúan diferente evaluaciones de manera diferente. Para este análisis se debe contar con dos objetos o personas a evaluar. Finalmente, podemos estudiar la varianza de las evaluaciones en un individuo dado. Cuanto menor sea la varianza mayores son las posibilidades de aparición del efecto halo.

Aparentemente ninguna de las medidas anteriores necesariamente señala a un efecto halo y ninguna de ellas nos permite distinguir entre un sesgo "de juicio" o un sesgo "de respuesta". Cuando las evaluaciones conciernen a dimensiones más o menos objetivas es posible comprobar la exactitud del evaluador y ver si las desviaciones de la exactitud efectúan en la dirección de la mayor concordancia, pero en otras situaciones, la noción de sesgo "de juicio" llega a ser más problemática, aunque si lo vemos como una tendencia más que como un error el concepto sigue siendo válido.

Aún queda el problema de determinar si el sesgo ocurre en el juicio o en la respuesta. Este solamente puede adscribirse mediante la comparación de los diferentes formatos de respuesta. A este respecto, no es suficiente comparar formatos de respuestas similares. Por ejemplo, diferentes escalas de evaluación pueden ser en gran parte lo mismo que otra, pero pueden diferir de forma dramática en otros formatos de medida.

Ha existido una tendencia a asumir que el efecto halo es un efecto "de juicio" (Nisbett y Wilson, 1977), pero esto no ha sido, hasta el momento, demostrado (West, 1983).

2. Indulgencia e inclemencia

La segunda forma de sesgo a considerar es la indulgencia y su complementaria la inclemencia. Como sus nombres indican son tendencias a realizar evaluaciones más positivas o negativas de las justificadas, es decir la tendencia a magnificar o minimizar las respuestas en las evaluaciones. De nuevo debe hacerse la distinción entre la indulgencia debida a los juicios de gravedad y la tendencia a dar respuestas más positivas o negativas dados aquellos juicios. Existen tres formas posibles de calibrar la magnitud de este sesgo (Nisbett y Wilson, 1977). Una forma consistiría en estudiar las diferencias entre las puntuaciones de los evaluadores para las mismas evaluaciones. En un análisis de varianza (ANOVA) esto se revelaría como un efecto principal del evaluador. Otro modo es comprobar el grado de desviación de la distribución normal, para un evaluador dado, una evaluación o de forma general. En el primer caso, el efecto principal del evaluador mostraría que, dadas las mismas evaluaciones, diferentes evaluadores dieron puntuaciones consistentemente más altas o más bajas. En el segundo caso podría verse que el grueso de las puntuaciones del evaluador estaba dirigida hacia un extremo de su distribución de respuestas. Al igual que en el efecto halo no podemos decir si realmente existe un sesgo erróneo en las respuestas a menos que dispongamos de otro criterio patrón más fiable con quien compararlo.

Si se encuentran efectos sistemáticos entre evaluadores podemos hablar de la existencia de sesgos en uno o más de los evaluadores, pero sin ser capaces de determinar cuales.

3. Tendencia Central-extrema

La tendencia central es una tendencia a realizar un número mayor de evaluaciones medias o neutrales de las justificadas.

La tendencia extrema es la tendencia a limitar las evaluaciones a los extremos de la escala. Podemos obtener un índice del grado de tendencia central o extrema comprobando las evaluaciones medias de cada individuo sobre un número de dimensiones. Si las medidas globales están cerca del punto medio de la escala podemos comprobar la varianza de las puntuaciones. También se puede comprobar la ausencia de un efecto principal de la evaluación o de lo evaluado para cada dimensión, que indicaría que los evaluadores no estaban discriminando entre los distintos evaluados. Finalmente, podemos valorar el grado de "Kurtosis" de las evaluaciones de una dimensión individual, si la media de las evaluaciones está alrededor del centro de la escala. La Kurtosis se refiere al aplanamiento de la curva de la distribución de las evaluaciones.

La tendencia central puede ser función del individuo, de la escala de evaluación o de la interacción entre ambos. Un individuo puede consistentemente ser supercauto e inclinarse a no desviarse demasiado de una posición central o moderada. Por otro lado, una escala particular puede, a causa de su irrelevancia y/o ambigüedad general, conducir a juicios no deseados por los evaluados. Finalmente, una dimensión o formato particular de evaluación puede ser más o menos relevante o ambiguo para diferentes evaluadores.

4. Restricción de rango y efectos frecuencia

La restricción de rango es, en cierta forma, similar a la tendencia central, pues es la tendencia a restringir las respuestas a una parte particular de la escala. Está indicada por la existencia de pequeñas varianzas en un evaluador en dimensiones individuales o por una alta Kurtosis. También puede conducir a una ausencia de un efecto principal de evaluado o de la evaluación (ANOVA) para una dimensión dada. Los efectos frecuencia fueron puestos de manifiesto por Parducci (1974) como un rasgo en las tareas de juicio perceptual. Consisten en la tendencia a usar todas las categorías de la escala en igual proporción esté o no justificado su uso, y por tanto es opuesto al sesgo de restricción de rango. Tanto la restricción de rango como los efectos frecuencia pueden ser producto del formato de respuesta ("efectos frecuencia particular"). Con un formato final abierto, por ejemplo, donde no existe una elección explícita de posicionamiento es difícil ver cuantos efectos frecuencia se manifestarían. Podríamos considerar la aquiescencia o tendencia a con-testar siempre verdadero o sí como un ejemplo particular de la respuesta sesgada de restricción de campo. Anastasi (1982) conceptualiza la aquiescencia como una variable continua, estando en extremo de la escala los que siempre responden que sí, y en el otro los que siempre responden que no.

5. Sesgo cambio de respuesta

Cronbach y Furbey (1970), y Howard y Dailey (1979) han señalado la importancia de este tipo de sesgo el cual es particularmente problemático para los experimentadores. Ocurre cuando una evaluación parece haber cambiado en el tiempo porque los puntos de comparación que determinan donde debería hacerse la evaluación también han cambiado. Por eso, un cambio en una evaluación después de exponerse a una manipulación experimental puede reflejar meramente un cambio en los puntos de comparación referenciales con los que se lleva a cabo la evaluación. Por ejemplo, un individuo que ha estado expuesto a una comunicación persuasiva que defiende una posición opuesta a su punto de vista, puede usar esa posición para juzgar su propio punto de vista cambiando por tanto la

evaluación que esa persona dio para expresar su punto de vista que en sí mismo ha permanecido inalterado. Este sesgo particular necesariamente se localiza en la respuesta más que en la evaluación porque asume que la evaluación permanece sin cambio.

6. Falsificación y deseabilidad o aceptabilidad social

Los inventarios autoadministrados están especialmente sujetos a la simulación o falsificación por parte de quien los cumplimenta. La mayoría de los ítems de dichos inventarios tienen una respuesta reconocible como socialmente más aceptable o deseable que las demás. En tales tests es posible que el que responda esté motivado a "falsificar bien" o elegir respuestas que puedan crear una impresión favorable de él, como cuando se solicita un empleo o se desea ser admitido en alguna institución educativa. Bajo otras circunstancias, el que responde puede estar motivado a "falsificar mal" intentando dar la impresión de estar más trastornado psicológicamente de los que está, como ocurre por ejemplo en el examen de individuos enjuiciados por causas criminales. Parece, por tanto, fácil crear deliberadamente la impresión que se desea en estos inventarios. Es interesante poner de manifiesto que también es posible llevar a cabo con éxito una simulación específica para un objetivo vocacional particular.

La existencia de una tendencia a elegir las respuestas socialmente deseables en un inventario autoadministrado no necesariamente indica un engaño o fraude deliberado por parte del que responde (Anastasi, 1982). Edwards (1957), que fue el primero en estudiar la deseabilidad social como variable, en principio la conceptualizó como "efecto fachada" o tendencia a "ponerse en el buen frente" de la cual el que responde es inconsciente. Esta tendencia puede indicar una ausencia de "insight" con respecto a sus propias características, un autoengaño, o una ausencia de voluntad para reconocer o hacer frente a sus propias limitaciones. Otros investigadores (Crowne y Marlowe, 1960) han presentado ciertas evidencias que sugieren que la fortaleza del conjunto de respuestas de deseabilidad social está relacionada con la necesidad más general del individuo de autoprotección, evitación de la crítica, conformidad social y aprobación social. Por otro lado, el individuo que elige los ítems desfavorables en una autodescripción puede estar motivado por la necesidad de atención, simpatía o ayuda en sus problemas personales. Las personas que buscan psicoterapia con frecuencia pretender parecer más trastornadas de lo que realmente están (Anastasi, 1982).

Varios procedimientos han sido desarrollados en un intento de enfrentarse al problema de la falsificación y respuestas sesgadas relacionadas. La construcción de ítems relativamente sutiles y neutrales socialmente puede reducir la operatividad de estos factores en algunos inventarios. En determinadas situaciones, las instrucciones del test y el establecimiento de una buena relación o entendimiento pueden motivar al examinando a responder francamente si logra ser convencido de que esto redundará en su propio beneficio. Sin embargo, este abordaje puede no ser efectivo en ciertas situaciones, y además no tendría mucho efecto en la respuesta si la deseabilidad social es inconsciente. Otras soluciones que se han intentado incluyen la verificación de claves para detectar la falsificación u otras respuestas sesgadas tales como la aplicación de la escala F del MMPI, y de términos correctores tales como los suministrados por la escala K del mismo inventario. Existe otro procedimiento, más dirigido a la prevención de la disimulación que a la detección de la misma, que consiste en el uso de ítems de

respuesta o elección forzada, que esencialmente se concreta en que el individuo a evaluar elija entre dos términos descriptivos o frases que parezcan tener una misma aceptabilidad pero que difieran en su validez.

7. Otras respuestas sesgadas

Otros posibles sesgos son el sesgo izquierdo-derecha (Payne, 1951), el error de contraste (Murray, 1938), y la aprehensión evaluada (Rosenberg, 1965). El sesgo izquierda-derecha habla por sí mismo y debe ser ubicado en la respuesta. El error de contraste se refiere a la tendencia para usarse a sí mismo como comparación para la realización de juicios sobre otros, y parecería, con bastante probabilidad, un error de juicio. La evaluación de la aprehensión se refiere a la presión que el experimentador ejerce sobre el que responde para que este dé la respuesta que debería dar o la que el experimentador desea, más que mostrar lo que los evaluados realmente piensen o sienten. No es estrictamente un sesgo de juicio, porque sus juicios pueden ser cabales y veraces, ni tampoco es un sesgo de respuesta en el mismo sentido en que otros describieron en que podía esperarse que fuera independiente de la forma de respuesta utilizada. Sin embargo, es una importante fuente de error en especial en experimentadores que necesitan dar pasos para asegurarse de que no comunican a los individuos las respuestas que ellos buscan. Otra respuesta sesgada sería la desviación o tendencia a dar respuestas poco frecuentes o inusuales.

Es necesario realizar una serie de puntualizaciones con respecto a los párrafos precedentes. En primer lugar, pocos de ellos tienen que ver de forma exclusiva con el lado de la respuesta del proceso. En segundo lugar, y surgiendo del primero, no existe, a priori, razón alguna para considerar que los sesgos están limitados a las evaluaciones, pues pueden surgir fácilmente en otros tipos de respuestas. Tercero, no existe una razón clara para considerar que en muchos casos disminuyen la validez o la fiabilidad de las evaluaciones como medida de actitud porque sus efectos se esperará que sean sistemáticos para un individuo o situación dados.

Los sesgos en las escalas de valoración no se limitan a los de los cuestionarios autoadministrados y parece conveniente,

llegado este momento, el esbozar de forma concisa las deficiencias que presentan las escalas de valoración cumplimentadas por un observador independiente. Aunque sus sesgos son menos obvios que los anteriormente comentados son tan graves como aquellos si no son oportunamente reconocidos (Snatih, 1981). La principal desventaja es la del sesgo del evaluador en el cual sus evaluaciones están influidas por su expectativa general de como debería estar de enfermo el paciente. Por ejemplo, generalmente se asume que los pacientes están más enfermos antes de comenzar el tratamiento que al finalizar el mismo. Una segunda desventaja es que el observador está influido en sus puntuaciones de gravedad por su experiencia general en pacientes que sufran de ese trastorno, siendo entonces una valoración relativa, pues su experiencia puede estar limitada a casos leves en la comunidad, o bien puede tener una gran experiencia en casos graves en el contexto hospitalario.

La investigación sobre respuestas sesgadas tales como la deseabilidad y aceptabilidad social, la aquiescencia o la desviación ha pasado por dos etapas principales. Cuando al principio fueron identificadas se las relacionó con una fuente de irrelevancias y de errores que debía ser eliminada de las puntuaciones de los tests, llevándose a cabo múltiples esfuerzos para eliminar su influencia a través de la reformulación de los ítems, el desarrollo de claves especiales o la aplicación de términos correctores. Más tarde, estas respuestas sesgadas fueron consideradas como características de personalidad amplias y duraderas que exigían una medición propia. En esta etapa fueron descritos los "estilos de respuesta". Anastasi (1982) ha puesto de manifiesto cierta asociación entre la respuesta sesgada de deseabilidad social y cierta tendencia de la conducta general del individuo para la búsqueda de aprobación social.

A pesar de las deficiencias referidas, el método autoadministrado es una forma útil de medida clínica y, como Nunnally (1978, p. 141) afirma "aun cuando los inventarios autoadministrados definitivamente tienen sus problemas al abordar la medida de las características de personalidad, actitudes, valores y una variedad de otros rasgos no cognitivos, representan con mucho el mejor abordaje posible". Además, tras diversos trabajos cuidadosos al respecto (Rorer, 1965; Fiske, 1971) han surgido serias dudas sobre el tamaño y persistencia de estos efectos.

BIBLIOGRAFIA

- Anastasi, A. Self-Report Inventories. En: Psychological Testing. Fifth Edition, MacMillan Publishing Co. Inc. London 1982.
- Cronbach LJ, Furby JL. How Should We Measure Change or Should Be. Psychological Bulletin 1970; 74: 68-86.
- Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. J Consul Psychol 1960; 29: 349-354.
- Derogatis LR, Covi L, Lipman RS, Davis DM, Rickels K. Social class and race as mediator variables in neurotic symptomatology. Arch Gen Psychiatry 1971; 25: 31-40.
- Derogatis LR, Lipman RS, Rickels K, Unlenhuth EH, Covi L. The hopkins symptom checklist (HSCL). A self-report symptom inventory. Behavioral Science 1974; 19: 1-15.
- Derogatis LR, Melisaratos N. The brief symptom inventory. An introductory report. Psychological Medicine 1983; 13: 595-605.
- Edwards AL. The social desirability variable in personality assessment and research. New York 1957.
- Fiske DW. Measuring the concepts of personality. Chicago 1971.
- Fowler RD. Automation and the Computer. In: MMPI Research Developments and Clinical Applications. Butcher (ed.) McGraw Hill. New York 1969.
- Howard GS, Dailey PR. Response Shift-Bias: A source of contamination of self-report measures. Journal of Applied Psychology 1979; 64: 144-150.
- Lanyon RI. Technological approach to the improvement of decision making in mental health services. J Consul Clin Psychol 1972; 39: 43-48.
- Less-Halley PR. Malingering emotional distress on the SCL-90-R. Toxic exposure and cancerphobia. Psychological Reports 1989; 65: 1203-1208.
- Murray HA. Explorations in personality. New York. Oxford University Press 1938.
- Nisbett RE, Wilson JD. The halo effect. Evidence for unconscious alteration of judgments. J Pers Soc Psychology 1977; 35: 250-256.
- Nunnally JC. An overview of psychological measurement. In: Clinical diagnosis of mental disorders. A hand book. Wolman (ed.) Plenum Press. New York 1978.
- Parducci A. Contextual effects: A range frequency analysis. En: Handbook of Perception. Cartarette & Friedman eds. New York

- Academic Press 1974; Vol. 2.
17. Payne SL. The art of asking questions. Princeton, New Jersey. Princeton University Press 1951.
 18. Platt JJ, Steer RA, Ranieri WF. Differences in the symptom check list-90. Profiles of black and white methadone patients. J Clin Psychol 1989; 45 (2): 342-345.
 19. Rorer LG. The great response-style Myth. Psychological Bulletin 1965; 63: 129-156.
 20. Rosenberg MJ. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. J Pers Soc Psychology 1965; 1: 28-42.
 21. Saal FE, Downet RG, Lahey MA. Rating the ratings: Assessing the psychometric quality of rating data. Psychological Bulletin 1980; 88: 413-428.
 22. Skilbeck WM, Acosta FX, Yamamoto J, Evans LA. Self-reported psychiatric symptoms among black, hispanic and white outpatients. J Clin Psychol 1984; 40(5): 1184-1189.
 23. Snaith RP. Rating scales. Br J Psychiatry 1981; 138: 512-514.
 24. Trorndike EL. A constant error in psychological ratings. Journal of Applied Psychology 1920; 4: 25-29.
 25. West RJ. Self-report measures in psychological experiments. Rating versus natural language measure. PhD (Sci) University of London 1983.
 26. Wilson JH, Taylor PJ, Robertson G. The validity of the SCL-90 in a sample of British men remanded to prison for psychiatric reports. Br J Psychiatry 1985; 147: 400-403.
 27. Woodworth RS. Personal data sheet. Stoelting Chicago 1918.
 28. Yamamoto J, Satele A, Fairbanks L, Samuelu R. Comparison of samoans in Samoa and California. Arch Gen Psychiatry 1981; 38: 231.