

SESGO DE GENERO EN LA MEDICION DEL NEUROTICISMO

Carmen Delgado Alvarez

Resumen

Las investigaciones sobre las escalas de neuroticismo (N) de Eysenck confirman invariablemente en diferentes medios culturales, que las mujeres obtienen puntuación significativamente más altas que los hombres. Se puede inferir a partir del resultado de la investigación, de la cual da cuenta este artículo, que el mayor neuroticismo de las mujeres parece ser únicamente un efecto del sesgo del instrumento de medida y no una característica del género femenino como se ha venido planteando hasta ahora. Esto debería afectar la investigación psicológica al menos en dos aspectos: 1) científico: iniciando una línea de investigación rigurosa dirigida a revisar toda la conceptualización psicológica del neuroticismo, y 2) ético: desideologizando desde una perspectiva de género, la conceptualización del neuroticismo.

Abstract

Research about Eysenck's Neurosis (N) Scales undoubtedly confirm, in different cultural environments, that women obtain a significant higher punctuations in contrast to men. From the result of this investigation could be infered that neurosis in women seem to be only an effect of slant from measurement instrument, instead of a characteristic of female gender as it has been outlined at present. It affects psychologic research in at least two aspects: 1. Scientific: starting investigations directed to review all psychologic conceptualization about neurosis, and 2. Ethic: cease ideology, from a gender perspective, related to the conceptualization of neurosis.

INTRODUCCION

Asistimos a una penetración de la perspectiva de género en el ámbito de la ciencia, quizás más lenta de lo que quisiéramos, pero persistente y eficaz desde mi punto de vista. Este avanzar costoso, pero eficaz, debería ser motivo de congratulación, no sólo para quie-

nes están en esta lucha, sino para todo científico y científica, de cualquier área del conocimiento, pues si el amor a la ciencia es el amor a la verdad, la perspectiva de género ha venido a desideologizar también la ciencia, entendiendo el término desideologizar en su acepción genuina de desencubrimiento del falso conocimiento al servicio de los intereses de quienes producen este falso conocimiento.

Sin embargo, no siempre funciona la lógica, y el intento de romper los muros con que se ha construido todo el edificio estructural de la civilización, provoca reacciones más ideológicas que científicas; pero casi siempre disfrazadas, o por lo menos revestidas, de ciencia. De este modo, la resistencia al cambio de perspectivas se vuelve más impenetrable y más difícil de combatir. ¿Qué decir cuando se nos objeta que nuestras posiciones, o las conclusiones de nuestras investigaciones, carecen del rigor científico requerido? Entramos en la guerra de paradigmas, en la que nadie puede ganar, porque acaba siendo una cuestión de "creencias" o adhesiones, sin criterios normativos universales. Y quedamos en la posición en la que estábamos. Desde mi punto de vista, éste es el problema del paradigma cualitativo que tantas aportaciones está haciendo desde la perspectiva del género. Al final, el cuestionamiento de su legitimidad científica, acaba siendo el punto final de todo debate; un punto infranqueable, porque no hay modo posible de resolver con criterios externos a las opciones epistemológicas, que a fin de cuentas acaban siendo *creencias*. Por esta razón considero que es importante plantear la batalla también en el terreno del contrincante: salir de esa *tierra de nadie*, donde cada quien puede quedarse donde estaba, porque todo acaba siendo cosa de opciones paradigmáticas. Con este propósito, este trabajo pretende ser una contribución a la tarea investigadora de la perspectiva de género, desde el paradigma cuantitativo, empirista, irrefutable por tanto para quienes se atrincheran en él.

¿También son sexistas los tests? Hay muchos modos de responder a esta pregunta, y uno de ellos es utilizando los mismos métodos de quienes defienden su neutralidad. Si alguna contribución pretendo con este trabajo, es ésta precisamente: contestar con sus mismos métodos a quienes sostienen, ingenua o ideológicamente, su neutralidad.

EL MODELO DE PERSONALIDAD DE EYSENCK

El estudio de la personalidad ha sido, y continúa siendo en Psicología, un punto de controversia y debate que enfrenta a los autores de diferentes enfoques psicológicos, quizás

más que en ninguna otra área de la psicología. Así, autores de los modelos psicodinámico, psicométrico, funcional, biológico, psicosocial, humanístico... compiten entre sí por demostrar la cientificidad de sus presupuestos, y la desacreditación de las teorías contrapuestas: si bien es cierto, que no todos con el mismo rigor científico.

La teoría de la personalidad de Eysenck pertenece al modelo psicométrico, y es una de las teorías que más ha impactado la investigación en psicología de la personalidad. Eysenck en su primera formulación identifica dos factores básicos de la personalidad: Neuroticismo y Extraversión; posteriormente añade un tercer factor: Psicoticismo (Eysenck, 1947; 1952; 1960; 1967; 1968). Estos factores ortogonales, y por tanto independientes entre sí, configurarían la personalidad de cualquier individuo. Cada factor se distribuye a lo largo de un continuum bipolar, que va desde la mínima hasta la máxima saturación, en el cual se situaría cada individuo. Así, la combinación de estos tres factores determinará la personalidad de cada sujeto, según su grado de Extraversión, Neuroticismo y Psicoticismo. Estos tres factores parecen ser universales, por cuanto han sido identificados en más de 40 países de culturas tan diferentes como Inglaterra, Uganda, USA, Sri Lanka, Tailandia, Singapur, China, Canadá, Hawai, Brasil, Holanda, Hong Kong, España, Yugoslavia, Zimbabwe, Puerto Rico, Holanda.

El modelo de Eysenck, uno de los clásicos en Psicología de la Personalidad, tiene una estructura jerárquica que parte de las conductas observadas en los individuos. Este punto de partida, le da una posibilidad de contrastación empírica que lo convierte en uno de los más fuertes epistemológicamente. Eysenck distingue en su modelo cuatro niveles en la conceptualización de la personalidad que se superponen en forma piramidal:

- * Nivel 1: Actos. Son las conductas observables que realizan las personas, como iniciar una conversación con una persona extraña.
- * Nivel 2: Conductas habituales. Son conductas que se repiten teniendo una alta probabilidad de aparecer en una persona determinada. Por tanto, vendrían a ser los actos que se repiten.

- * Nivel 3: Rasgos, Serían conjuntos de conductas que aparecen asociadas. Así por ejemplo, cuando aparecen conductas habituales tales como iniciar conversaciones con otra gente, estar frecuentemente acompañado, ir frecuentemente a fiestas,... se puede postular un rasgo de sociabilidad.
- * Nivel 4: Tipos. Serían conjuntos de rasgos. Así, cuando la sociabilidad va asociada a actividad física, apertura, variabilidad, ... se puede hablar de una dimensión de personalidad o tipo.

La personalidad estaría organizada por tanto, de una manera jerárquica: desde lo más específico que hacen las personas (nivel 1) hasta lo más general o abstracto, que es su configuración en rasgos: N,E, y P (nivel 4). Eysenck, Barret, Wilson & Jackson (1992) presentan del siguiente modo, la caracterización de un sujeto que tuviera puntuaciones altas en cada uno de los factores:

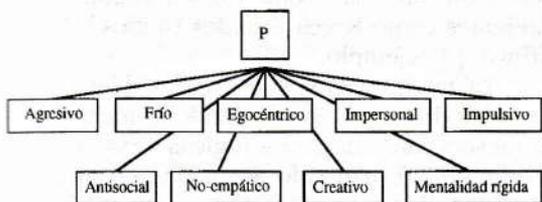
* Neuroticismo (N):



* Extraversión (E):



* Psicoticismo (P):



Epistemológicamente, este modelo ofrece al menos tres ventajas, según el propio Eysenck (1991):

- 1) El modelo incorpora la distinción entre estados y rasgos, originalmente desarrollada por Cicerón hace unos 2000 años. El nivel más bajo, el uno, sería una medida del estado de una persona; el nivel más alto, sería una medida de un rasgo.
- 2) El modelo incorpora los principios de agregación (Rushton, Brainerd & Presley, 1983), según los cuales los conceptos del nivel dos se construyen sobre los del nivel uno, y así sucesivamente.
- 3) El modelo es empíricamente contrastable, y ha producido un gran número de tests con este propósito.

Sin pretender en este artículo hacer una exposición sucinta de la teoría de personalidad de Eysenck, se puede concluir con todo rigor que es uno de los modelos de mayor reconocimiento y prestigio científico en psicología de la personalidad, y de los que mayor número de investigaciones ha generado y sigue generando, tanto para confirmarlo como para refutarlo. Las técnicas psicométricas de evaluación de la personalidad que se derivan de él (EPI, EPQ, JEPI, EPQJ,...) son asimismo ampliamente utilizadas, especialmente en situaciones de selección: admisión a colleges, selección de personal, etc...

DIFERENCIAS DE GENERO EN NEUROTICISMO

Eysenck (1978) precisa la significación del término Neuroticismo en su teoría y le da el significado de comportamientos asociados a emociones fuertes-intensas no adaptativas, ante algo ilógico, absurdo o irrelevante. De un modo más científico Eysenck & Eysenck (1975) en el manual de Eysenck Personality Questionnaire, definen una persona típicamente neurótica como ansiosa, inquieta, melancólica y frecuentemente deprimida, probablemente con problemas de sueño y desórdenes psicosomáticos; demasiado emocional, con fuertes reacciones a estímulos pequeños, y dificultades para volver al equilibrio después de cada experiencia emocionalmente excitante; sus fuertes reacciones emocionales, a veces

irracional y rígidas, interfieren con su ajuste. Eysenck & Gudjonsson (1989) caracterizan a la persona neurótica como ansiosa, tensa, irracional, tímida, malhumorada, emocional, con sentimientos de culpa y con baja autoestima.

Un sujeto neurótico sería un sujeto ansioso, deprimido, con sentimientos de culpa, baja autoestima, tenso, irracional, tímido, melancólico y emocional. Lógicamente, cada uno de estos rasgos, los poseería en una cantidad determinada, y en cada uno de ellos se situaría en un punto de *continuum* que los cuantifica. La combinación de estos elementos, determinaría la "cantidad" de Neuroticismo de un sujeto. Eysenck precisa que su concepto de Neuroticismo, no es un concepto patológico; cuando habla de Neuroticismo habla de una dimensión presente en personalidades normales, y lo define como predisposición a reacciones neuróticas, sin que ello signifique que las conductas de las personas sean patológicas.

No obstante en esta conceptualización del Neuroticismo, parece existir un acuerdo general entre las diferentes corrientes psicológicas, en que las conductas que caracterizan a las personas neuróticas están relacionadas con características socialmente no deseables. De hecho, en situaciones de selección, la puntuación en Neuroticismo se utiliza para rechazar a los sujetos más neuróticos, no para elegirlos. No es una dimensión "neutra" como pretende Eysenck cuando se molesta en insistir en que su Neuroticismo, no es una dimensión patológica de la personalidad. La no deseabilidad de esta dimensión, queda corroborada en la investigación de Furnham (1984), cuando encuentra que los sujetos tienden a declararse a sí mismos menos neuróticos de lo que reflejan sus puntuaciones en Neuroticismo. Dunnett, Koun & Barber (1981) en un estudio con los ítems del EPI encuentran que mientras la proporción media de ítems de Extraversión va en la dirección de la deseabilidad social tanto para hombres como para mujeres, la proporción media de ítems de Neuroticismo va significativamente en la dirección de la indeseabilidad. Otros estudios de distinto tipo también ilustran la relación inversa entre deseabilidad social y expresión de características tipificadas como neuróticas (Velicer & Weiner, 1975; Cowles, Darling & Skanes, 1992; Levin & Montag, 1987; Lodhi & Thomas, 1991). Rim

(1982) describe a los sujetos con puntuaciones altas en Neuroticismo como más realizadores de conducta expresiva (expresión de emociones), y más conscientes de las constricciones de las normas sociales externas. Numerosos estudios relatados por Francis (1993), señalan correlaciones del Neuroticismo con características no deseables, incluyendo patologías y disfunciones físicas como enfermedad coronaria, bulimia, cáncer, hipocondriasis, trastornos de la respuesta sexual,... Así pues, por sí el sentido común pareciera poco riguroso como criterio científico, queda suficientemente documentada la indeseabilidad social asociada a la dimensión de Neuroticismo, por más que Eysenck insista en la normalidad de esta dimensión.

Invariablemente, con las diferentes y sucesivas escalas de Neuroticismo de Eysenck, las mujeres obtienen puntuaciones más altas que los hombres. En población adulta se encontraron estas diferencias en Alemania, Australia, Bangladesh, Brasil, Bulgaria, Canadá, China, Egipto, El Salvador, España, Francia, Grecia, Groenlandia, Hawai, Holanda, Hong Kong, Hungría, India, Irán, Israel, Italia, Japón, Korea, Nigeria, Polonia, Rumanía, Rusia, Sicilia, Singapur, Sri Lanka, Tailandia, Uganda, USA, y Yugoslavia. En población infantil y jóvenes, se encontraron diferencias sexuales en Australia, Austria, Canadá, China, Dinamarca, Egipto, El Salvador, España, Grecia, Groenlandia, Hawai, Holanda, Hong Kong, Hungría, India, Irlanda, Japón, Nueva Zelanda, Puerto Rico, Singapur, Sri Lanka, Suecia, Uganda, USA, Yugoslavia y Zimbabwe (Francis, 1993; Delgado, 1994). La inmediata conclusión tras lo expuesto anteriormente, muy lógica por lo demás, es que no cabe duda de que la personalidad femenina es menos deseable y se ajusta menos a las condiciones requeridas para el éxito social. Y esto como fenómeno casi universal, pues estas diferencias no se circunscriben a un ámbito cultural determinado, sino que aparecen una y otra vez en culturas tan diferentes como Korea, Estados Unidos y Zimbabwe, por ejemplo.

La interpretación fácil de las diferencias entre los dos sexos, convirtiéndolas incluso en supuestas diferencias biológicas (Eysenck & Wilson, 1979), han sido aceptadas sin apenas cuestionamiento posible, dada la *cientificidad*

de sus presupuestos. Por algo se ha defendido la incuestionabilidad de la *neutralidad científica* y más aún tratándose de datos empíricos, contrastables y cuantificables. Pero una grieta comienza a abrirse cada vez con más fuerza en el cuestionamiento de los instrumentos de medida, y lo interesante de estos cuestionamientos es que se sostienen *con las mismas armas* del empirismo, con lo cual la contraargumentación de desacreditar el paradigma científico desde el cual se critica, como se suele hacer casi siempre no vale en este caso.

Algunos estudios, empiezan a sugerir que estas diferencias en Neuroticismo, quizás no estén mostrando diferencias reales entre hombres y mujeres en esta dimensión, sino más bien pareciera ser efecto del mismo instrumento de medida, que resulta "sospechoso" de estar sesgado (Jorm, 1987; Francis, 1993). Es cierto que no se ha mostrado todavía suficiente evidencia empírica para resolver la disyuntiva con argumentos sólidos. Los defensores de las diferencias reales describen a las mujeres como más emotivas que los hombres y con mayor tendencia a padecer problemas neuróticos de una u otra clase: ansiedades, preocupaciones, depresión, etc... y en algunos casos llegan a sostener que los roles masculinos y femeninos están profundamente incorporados a nuestra naturaleza biológica (Eysenck & Wilson, 1979); aunque este punto de vista está en contra de la mayoría de investigadores de las diferencias sexuales (Kohlberg, 1966; Mischel, 1970; Maccoby, 1974, 1983; Unger, 1979; Bem, 1981; 1983; López, 1984). Resulta cuanto menos llamativa la siguiente afirmación de Eysenck y Wilson (1979, p.63), tras señalar la tendencia de las mujeres a padecer problemas neuróticos:

Sin lugar a dudas la mayoría de las mujeres parecen encantadas con su rol femenino, igual que la mayoría de los hombres están encantados con el suyo. Que sea deseable privar a las personas de esta felicidad con la incierta esperanza de que podrán hallar alguna satisfacción en algún tipo de cultura y de roles unisexuales nos parece algo muy dudoso.

No resulta fácil entender esa "felicidad neurótica" de las mujeres, a la que parecen re-

ferirse los autores, y que parece situar las interpretaciones de las diferencias sexuales en Neuroticismo más en un terreno ideológico, que en comprobaciones estrictamente científicas. En este estado de cosas parece más verosímil la opinión de Jorm (1987), quien considera que la explicación más simple de estas diferencias podría ser la presencia de sesgo sexual en las escalas de Neuroticismo, ya que el Neuroticismo puede manifestarse de forma diferente en hombres y mujeres. No es de extrañar que los diferentes procesos de socialización sufridos por hombres y mujeres, determinen expresiones diferentes de tendencias igualmente "neuróticas" y que, como señala Jorm (1987), sea más probable que alguien lllore más fácilmente si es mujer que si es hombre. Si las escalas de Neuroticismo se construyen conteniendo más ítems apropiados para las mujeres, las diferencias de género que aparecieran en las puntuaciones medias no serían más que la expresión del sesgo en la selección de los ítems.

La investigación generada en torno a este problema, apunta hacia la existencia de dos estilos de Neuroticismo en el Eysenck Personality Inventory, Roger & Nesselhoever (1987) los han denominado *hypochondriasis y social sensitivity*, pero no han estudiado la relación de estos estilos con el género. Francis (1993) estudiando los ítems del Junior Eysenck Personality Inventory (JEPJ), Junior Eysenck Personality Questionnaire (JEPQ), Eysenck Personality Inventory (EPI), y Revised Eysenck Personality Questionnaire (REPQ), encuentra una dimensión ligada al sexo a la que denomina NEURO-S y otra dimensión libre del sexo a la que llama NEURO-A. Estas dos dimensiones no sólo difieren entre sí en su correlación con el sexo, sino que parecen medir aspectos diferentes del constructo Neuroticismo, por cuanto se relacionan de forma diferente con las escalas Extraversión (E), Psicoticismo (P) y Sinceridad (L), y estos hallazgos son confirmados en medios culturales diferentes: Canadá, USA y Reino Unido. Estos resultados parecen apoyar más la hipótesis de presencia de sesgo en el instrumento de medida, que diferencias reales entre los sexos, puesto que en la dimensión NEURO-A no aparecen las supuestas diferencias de género. Pero estos resultados, como la misma autora sugiere, no son suficientes para

inferir sesgo en las escalas; se requiere mayor investigación, y más específica de sesgo, para llegar a la conclusión de si las diferencias sexuales en N representan un fenómeno real o se deben a un efecto de sesgo de género en el instrumento. Por esta razón, en esta investigación se hizo un estudio del sesgo de los ítems con una de las técnicas psicométricas específicas para su detección: el método TID.

EL SESGO EN LOS TESTS PSICOLOGICOS

Jensen (1980) constata la presencia de diferencias sexuales en situaciones de selección, en las que se utilizan los tests mentales para tomar decisiones de admisión: programas de escuela especial para sujetos con retardo o para superdotados, admisiones a colleges, selección de personal... La existencia de esta diferencia entre hombres y mujeres (como en el caso de las diferencias raciales o entre clases sociales), plantea la necesidad ética y científica de determinar con rigor si estas diferencias mostradas entre hombres y mujeres son reales, o bien son diferencias artificiales debidas al sesgo de los tests.

La presencia de sesgo en los tests, parece ser algo inevitable por el mismo proceso de construcción, hasta tal punto que algunos autores afirman que estrictamente no existen pruebas completamente exentas de sesgo. Pero aún así, la tarea psicométrica sería tratar de detectar la cantidad de sesgo "tolerable". El método más eficiente para evitar en lo posible el sesgo de los tests parece ser el de asegurar la validez de contenido de los ítems que lo integrarán. Una vez filtrado este elemento, existen análisis estadísticos que permiten detectar el sesgo que aún pudiera persistir en los ítems.

Si bien el sesgo sexual en tests de inteligencia ha sido investigado en campos de aplicación muy concretos (selección de colleges y selección de personal principalmente), y muchas veces a instancias de demandas judiciales por parte de grupos étnicos discriminados con su uso, no ocurre así con variables de personalidad en las que apenas se ha entrado. No he encontrado en ninguno de los principales manuales sobre sesgo de los últimos años (Jensen, 1980; Berk, 1982; Osterlind, 1983;

Reynolds & Brown, 1984) ninguna referencia a esta cuestión, salvo la referencia al interés de que sea investigado. Sin embargo, cuestionarios de personalidad ampliamente utilizados en medios culturales muy diferentes arrojan consistentemente diferencias entre grupos definidos. En Neuroticismo, se confirman insistentemente diferencias en función de la clase social, siendo *lógicamente* la clase baja la que puntúa más alto; y diferencias en función del sexo, a las que se acaba de hacer referencia. La forma general de interpretar estos resultados, es concluir que en la clase social baja hay más Neuroticismo que en la clase alta, y que las mujeres son más neuróticas que los hombres. Ha sido la discusión social y teórica, ajena en gran parte al círculo psicométrico especializado, la que ha introducido la necesidad de investigar sobre la presencia de sesgo en los instrumentos de medida, antes de afirmar la diferencia entre los grupos.

Se define el sesgo como un error sistemático en el proceso de medida, en virtud del cual todas las mediciones obtenidas quedan afectadas por exceso, *sobreestimación*, o por defecto, *subestimación* (Osterlind, 1983). Conceptualmente, el sesgo es algo distinto de los conceptos de justicia, igualdad, prejuicio, o cualquier otra connotación asociada con su uso en el lenguaje común. En sentido técnico, el sesgo denota la distorsión consistente de un estadístico. Sin embargo, es obvio que de la investigación estrictamente técnica de la presencia de sesgo en un test, se derivan consecuencias sociales directamente relacionadas con los conceptos de justicia e igualdad a los que nos referíamos.

El sesgo de los ítems, y por ende de los tests, aparece a partir de los años setenta en la literatura psicométrica especializada. Los textos clásicos (Lord & Novick, 1968; Thorndike, 1971) prácticamente no lo citan. Tras el polémico artículo de Jensen (1969) este problema se introduce en el ámbito de la investigación. A partir de los años ochenta, aparecen las publicaciones de Shepard, Camilli & Averill (1981), Osterlind (1983), Reynolds & Brown (1984), y el tema del sesgo comienza a aparecer como un capítulo importante en las publicaciones psicométricas (Lord, 1980; Crocker & Algina, 1986; Hambleton, 1989; Hambleton, Swaminathan & Rogers, 1991). Desde el punto

de vista psicométrico, se considera que un test está insesgado si todos los individuos que tienen el mismo nivel de habilidad en la variable medida, tienen igual probabilidad de responder correctamente a los ítems del test, sin tener en cuenta a qué subgrupo pertenecen (Pine, 1977). En este caso, podríamos aceptar como libres de sesgo los ítems de la escala N de Eysenck, si sujetos con igual nivel de Neuroticismo en la escala tienen igual probabilidad de dar respuesta neurótica al ítem, independientemente del género. Si, por el contrario, ante un ítem determinado las mujeres presentan una probabilidad mayor que los hombres de dar una respuesta neurótica al ítem, y ambos -hombres y mujeres- tienen el mismo nivel de Neuroticismo, se puede inferir que ese ítem está sesgado en contra de las mujeres. Es preciso subrayar que no es correcto inferir sesgo por el mero hecho de que se obtengan puntuaciones más altas en un grupo, ya que esas diferencias pueden recoger diferencias reales. Para que se pueda inferir sesgo, es preciso que los sujetos presentando *igual nivel de Neuroticismo*, presenten probabilidad de respuesta diferente ante un ítem por el hecho de *pertenecer a un grupo tipificado*; en este caso, por el hecho de ser mujer.

Las fuentes de sesgo en los tests psicológicos son numerosas, y vienen generadas principalmente por el distinto bagaje cultural, social, económico, etc... de los sujetos, o en términos más conductuales por su historia estimular. Pero, como señala Muñiz (1990), dado que estos antecedentes históricos de los sujetos nunca serán los mismos, y pueden ser marcadamente distintos según la subcultura, si un test se apoya más en la subcultura de unos que en la de otros, tendrá altísimas probabilidades de no ser equitativo, de estar sesgado. El problema puede tener serias repercusiones sociales si es precisamente una de las dos subculturas, obviamente la dominante, la que construye los tests para todos. La cuestión de sesgo sexual parte de la consideración de que niños y niñas, desde temprana edad, son socializados en roles tradicionalmente femeninos y masculinos (diferencias en juegos, juguetes, experiencias, oportunidades de aprendizaje). Los contenidos de los ítems ciertamente no son independientes de estas diferencias sexuales en la experiencia, y algún test determinado

puede contener más ítems de signo femenino que masculino, o viceversa, resultando ser un test sesgado por los ítems que lo componen. Así como en inteligencia es más ventajoso obtener puntuaciones altas, en Neuroticismo ocurre lo contrario. En personalidad, si puntuaciones altas en Neuroticismo son socialmente indeseables y socialmente no-atractivas, la presencia de más ítems ligados al modo específico en que las mujeres manifiestan el Neuroticismo, podría ser una fuente de sesgo en su contra, que podría estar explicando esas diferencias sexuales que consistentemente aparecen en las investigaciones realizadas en países tan diferentes (Jorm, 1987). Ahora bien, es preciso insistir una vez más en que la mera aparición de diferencias entre dos grupos no significa necesariamente sesgo en contra del grupo con puntuación más desventajosa. Podría darse incluso la existencia de sesgo, pero en contra del grupo con puntuaciones más favorables, como en el caso investigado por Bickel, Hammel & O'Connell (1975).

Osterlind (1983) señala fundamentalmente cinco procedimientos de detección de sesgo, cada uno de los cuales representaría un rango de aproximaciones para descubrir el grado de sesgo que puede estar presente en unos ítems específicos. Cada uno de ellos representa una familia de rutas a seguir. No son estrategias independientes unas de otras, se superponen y se construyen unas sobre otras. Hambleton, Clauser, Mazor & Jones (1993) hacen una revisión de estas técnicas, en relación con los avances en la detección de sesgo.

En este trabajo se utilizará la metodología "Dificultad de los Ítems Transformada" (TID), también conocida como "método Delta". Se ha utilizado el paquete estadístico StatView 512+ de Macintosh, para los cálculos pertinentes (Martín & Delgado, 1993). Los estudios comparativos de la eficacia de los distintos métodos de detección de sesgo, parecen indicar que el método TID, junto con el método Ji cuadrado de Scheuneman (1975) es más eficaz cuando el sesgo afecta sólo al parámetro de dificultad (Burril, 1982). Este resultado es obtenido por Rudner, Getson & Knight (1979) cuando se estima el sesgo con datos simulados. Ironson & Subkoviak (1979) con datos empíricos, muestran un porcentaje de concordancia de resultados entre la técnica TID y

la técnica de tres parámetros de la Curva Característica del Ítem (ICC) del 33,3%; y entre TID y Ji cuadrado del 37,5%. Shepard, Camilli & Averill (1981) concluyen que el procedimiento TID de Angoff puede ser utilizado en lugar del modelo de Rasch para la detección del sesgo.

METODO

Se aplicó la forma B del Eysenck Personality Inventory (EPI), en la versión española de TEA Ediciones (Madrid, 1982) a una muestra de 940 sujetos que siguieron el curso de admisión en la Universidad Centroamericana "José Simeón Cañas" (UCA) de El Salvador en enero de 1994. La distribución por sexos fue la siguiente: 502 hombres y 438 mujeres. La media de edad resultó ser 18.533 años, con una desviación típica de 2.083. Considerada la edad por sexos, se obtuvo una media de 18.709 años y desviación típica de 2.32 para los hombres y media de 18.336 y desviación típica de 1.764 para las mujeres. Por tanto, se pueden considerar igualadas en edad y nivel cultural las muestras de hombres y mujeres.

La escala de Neuroticismo (N) está compuesta por 24 ítems de los 57 que componen el EPI: ítems N^{os} 2, 4, 7, 9, 11, 14, 16, 19, 21, 23, 26, 28, 31, 33, 35, 38, 40, 43, 45, 47, 50, 52, 55, y 57. Además de la escala N, este cuestionario incluye una escala de Extraversión (E) y otra escala de Insinceridad (L). En la versión española la escala L está formulada como escala de Sinceridad (S). Se estimó la consistencia interna de la escala N mediante alfa de Crombach obteniéndose un valor .7038 similar a los obtenidos en otras investigaciones.

RESULTADOS

Dimensiones de Neuroticismo y Correlación con el sexo

Las mujeres obtienen una media en neuroticismo de 11.575 frente a una media de 10.452 obtenida por los hombres. El contraste de medias dio un valor t de Student de 4.345, significativo para $p < 0.0001$. Por tanto, se confirman en El Salvador los resultados obtenidos

en otros países, que atribuyen a las mujeres puntuaciones más altas en Neuroticismo. La interpretación simple de este resultado sería concluir una vez más que las mujeres son más neuróticas que los hombres. Sin embargo, un primer análisis de los elementos que componen la escala, siguiendo el método correlacional utilizado por Francis (1993), revela la existencia de dos grupos de ítems que se comportan de modo diferente en su relación con el sexo de los sujetos que responden: 1) un set de ítems que muestran correlación significativa con el sexo, al que hemos llamado NEURO-S siguiendo la terminología de Francis; 2) un set de ítems que no correlacionan con el sexo, y que se comportan igual con hombres que con mujeres, al que llamamos NEURO-A.

La tabla 1 muestra las correlaciones con el sexo, obtenidas para cada uno de los ítems de las dos dimensiones NEURO-S y NEURO-A, que componen la escala N del EPI:

Tabla 1
Coeficientes de Correlación con el Sexo

NEURO-S		NEURO-A	
Item	r	Item	r
4	0.120	2	0.034
7	0.208	9	-0.006
28	0.111	14	0.016
38	0.139	16	-0.044
43	0.198	19	-0.020
45	0.145	21	-0.072
47	0.227	23	-0.011
57	0.171	26	-0.009
		31	0.058
11	-0.119	33	-0.023
		35	0.048
		40	0.040
		50	-0.009
		52	0.017
		55	0.020

Todos los ítems que componen la dimensión NEURO-S, excepto el n^o11, correlacionan positivamente con el sexo femenino. El ítem n^o11 tiene la siguiente formulación: "¿Ocurre con frecuencia que toma usted sus decisiones demasiado tarde?". Por tanto, ocho ítems de los nueve que componen la escala NEURO-S, dan puntuaciones superiores cuando el sujeto que responde es mujer.

Las mujeres dan puntuaciones superiores a los hombres cuando se comparan las puntuaciones de Neuroticismo; pero cuando se comparan separadamente las dimensiones NEURO-S y NEURO-A, estas diferencias sólo aparecen en NEURO-S, tal como se muestra en la tabla 2:

Tabla 2

Diferencias en NEURO-A y NEURO-S

	Hombres	Mujeres	t
Neuroticismo	10.452	11.575	4.345*
Neuro-A	6.468	6.447	0.114
Neuro-S	3.984	5.128	9.846*

* $p \leq 0.0001$

¿Qué decir a propósito de estos primeros resultados? En primer lugar, que la conceptualización del Neuroticismo parece incluir un componente de género identificable; y en segundo lugar, que este componente resulta claramente desventajoso para las mujeres. Sin embargo, rigurosamente hablando no se puede afirmar todavía la presencia de sesgo. Es posible que este componente ligado al sexo, esté reflejando algo real; es posible que esté reflejando tendencias neuróticas, más presentes en las mujeres que los hombres. Por esta razón, se procederá a hacer un análisis del sesgo. Con este análisis se pretende demostrar, que con el mismo nivel de neuroticismo, las mujeres dan más respuestas neuróticas a los ítems que los hombres. Muñiz (1990) ejemplifica de un modo muy gráfico en qué consiste el sesgo: consideremos un metro de metal con el que pretendemos medir dos tuberías de agua. Supongamos que una de ellas transporta agua fría, y la otra agua caliente. Y supongamos que las dos miden exactamente un metro. El instrumento de medida, en este caso el metro, estará insesgado si al hacer las mediciones se obtiene en ambos casos una medición idéntica: un metro. Pero, ¿qué podría ocurrir con este instrumento? Por su composición de metal, podría dilatarse con el calor al contacto con la tubería de agua caliente, y entonces la medida que nos daría ya no sería de un metro, sino algo menos del metro. En este caso, estaríamos trabajando con un instrumento sesgado contra las tuberías de

agua caliente. Sería un instrumento inadecuado para comparar tuberías de agua fría con tuberías de agua caliente. Volviendo a la escala de neuroticismo, el hallazgo de la dimensión NEURO-S muestra que algunos componentes del instrumento de medida dan mediciones más altas cuando se miden mujeres que cuando se miden hombres; pero es preciso verificar si esto se debe al objeto medido (en este caso se trataría de diferencias reales entre hombres y mujeres), o a la naturaleza del instrumento (que distorsiona su medición por el componente de género que incluye en su composición).

ANÁLISIS DE SESGO

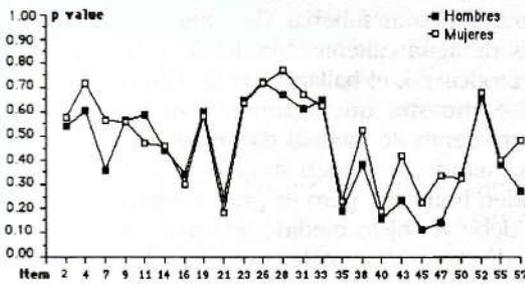
Siguiendo el método TID ya referido, se calculó la proporción de mujeres y de hombres que responden positivamente a cada ítem (p-value). El p-value es un índice de dificultad de ítem, que expresa de forma inversa cuán difícil resulta un ítem para un grupo. Así, un ítem ante el cual el 30% de las mujeres responden afirmativamente tiene un p-value para el grupo de mujeres de 0.30, y resulta más difícil para este grupo, que un ítem ante el cual el 80% de las mujeres responden afirmativamente. Es decir, que cuanto más bajo es el p-value de un grupo, más difícil es ese ítem para ese grupo. Por eso hablamos de índice de dificultad inverso. En este caso, cuanto menor sea el p-value de un grupo, menos probable es que los miembros de ese grupo den respuestas neuróticas ante el ítem. Desde el punto de vista social, esto se traduce en que lo más ventajoso es tener un p-value bajo.

Teóricamente, los ítems deberían tener el mismo p-value para hombres y para mujeres, ya que deberían resultar igualmente difíciles o fáciles para ambos grupos. Sin embargo, tal como se observa en el gráfico 1, en la mayoría de los ítems las mujeres tienen un p-value más alto que los hombres:

Los ítems 11, 16 y 21 son los únicos que dan mayor probabilidad de respuesta neurótica en los hombres frente a 10 ítems con mayor probabilidad de respuesta neurótica en las mujeres.

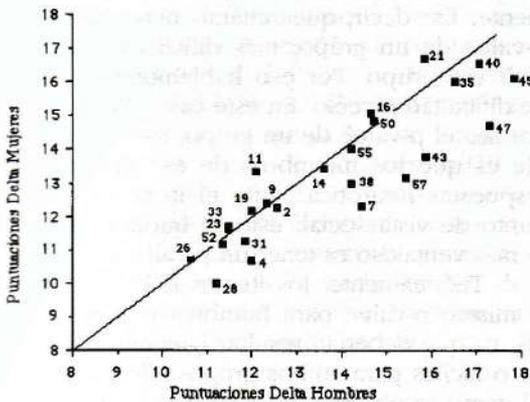
Se efectuó la transformación de p-values en puntuaciones delta (Δ), tomando Z como el percentil (1-p) de la distribución normal estandarizada (Osterlind, 1983). El procedimiento es

GRAFICO 1



el siguiente: puesto que el p-value es un índice de dificultad inverso, se calculó su complementario $q = 1-p$. Este valor "q" es un índice directo de dificultad; es decir, los valores altos de "q" representan índices altos de dificultad. Mediante las tablas de curva normal, se determinaron los valores Z correspondientes a cada valor "q". Estos valores Z se transformaron en puntuaciones de la escala típica derivada delta (Δ), de media = 13 y desviación típica = 4 ($\Delta = 13 + 4Z$). Obtenidas las puntuaciones delta correspondientes para cada ítem en cada grupo, se obtuvo la nube de puntos representada en el gráfico 2:

GRAFICO 2



Si los ítems tuvieran el mismo índice de dificultad para ambos grupos, los puntos que representan a los ítems se agruparían en torno al eje principal de 45 grados; pero aparece una clara tendencia a situarse por debajo de esta recta, lo cual indica que los hombres tienen menor probabilidad que las mujeres, de dar respuestas neuróticas a los ítems. La presencia de sesgo se cuantifica a través de la distancia del ítem al eje. Aunque algunos autores como Rudner & Convey (1978) estiman la dis-

tancia a este eje de 45 grados, Angoff & Ford (1973) recomiendan utilizar el eje mayor de la elipse para el cálculo de la distancia, puesto que este eje mayor es el que mejor representa la nube de puntos por cuanto minimiza la distancia de un set de ítems dado al eje principal. El objetivo consiste entonces en obtener la función de distancia de cada ítem; esto es, la distancia perpendicular de cada ítem al eje principal. La estimación de las distancias por el procedimiento referido de Angoff & Ford (1973), se efectuó mediante las fórmulas:

$$\text{El eje principal viene dado por: } Y = aX + b$$

donde:

$$a = \frac{S_y^2 - S_x^2 + \sqrt{(S_y^2 - S_x^2) + 4 r_{xy}^2 S_x^2 S_y^2}}{2 r_{xy} S_x S_y}$$

$$b = \bar{Y} - a \bar{X}$$

y el índice de distancia por:

$$d = \frac{a X_j - Y_j + b}{\sqrt{a^2 + 1}}$$

donde:

- X_j : Puntuación delta del ítem j en el grupo X
- Y_j : Puntuación delta del ítem j en el grupo Y
- a y b: vienen dados por las fórmulas citadas.

Se tomaron como índices de sesgo distancias superiores a ± 0.75 puntuaciones Z (Osterlind, 1983).

En la tabla 3 se presentan las distancias de cada ítem al eje principal. Las distancias con signo negativo, corresponden a los ítems con mayor probabilidad de respuesta neurótica en las mujeres, por tanto, sospechosos de estar sesgados contra las mujeres. Por el contrario, las distancias positivas corresponden a los ítems con mayor probabilidad de respuesta neurótica en los hombres. Cuando estas

distancias superan el valor 0.75 se puede considerar que el ítem está sesgado; el signo de la distancia indica contra qué grupo actúa el sesgo. Así se obtiene sesgo en los siguientes ítems:

- a) Contra los hombres: 11, 16, y 21.
 b) Contra las mujeres: 4, 7, 28, 43, 47, y 57.

Tabla 3
 Distancias al eje mayor

Item	D	Item	D
2	0.007	31	-0.323
4	-0.860 *	33	0.277
7	-1.105 *	35	0.468
9	0.282	38	-0.440
11	1.151**	40	0.580
14	0.294	43	-0.850 *
16	0.925**	45	-0.295
19	0.295	47	-1.038 *
21	1.424**	50	0.687
23	0.183	52	-0.070
26	0.021	55	0.356
28	-0.904 *	57	-1.059 *

* Indicativo de sesgo contra las mujeres

** Indicativo de sesgo contra los hombres

DISCUSION

El contenido de los ítems que han mostrado sesgo en contra de uno u otro sexo tienen la siguiente formulación:

- a) Contra los hombres:
 Item 11: ¿Ocurre con frecuencia que toma Usted sus decisiones demasiado tarde?
 Item 16: ¿Se siente con frecuencia "harto" de todo?
 Item 21: ¿Está Usted con frecuencia "en la luna?"
- b) Contra las mujeres:
 Item 4: ¿Se siente Usted unas veces triste y otras alegre, sin motivo?
 Item 7: ¿Es usted una persona caprichosa?
 Item 28: Cuando está disgustado, ¿necesita algún amigo para contárselo?
 Item 43: ¿Sufre Usted de los nervios?

Item 47: ¿Sufre Usted dolores de cabeza o jaquecas muy fuertes?

Item 57: ¿Siente a menudo molestias digestivas frente a un hecho o situación importante?

El sesgo contra las mujeres aparece en los ítems que hacen referencia a *somatizaciones* (43, 47, 57), y *conducta expresiva de emociones* (4, 7, 28). Mientras que el sesgo contra los hombres aparece en ítems que se refieren al modo de *enfrentar la realidad*: tomar decisiones demasiado tarde (11), sentirse "harto" de todo (16), o escapar de la realidad (21). Esto parece estar en consonancia con las dos dimensiones planteadas por Roger & Neshoever (1987) como *hypocondriasis* y *sensibilidad social*, y con la descripción que hace Rim (1982) de los sujetos neuróticos como realizadores de conducta expresiva y con mayor consciencia de las constricciones de las normas sociales externas. Lo que se puede añadir a partir del sesgo sexual detectado en la escala, es que cada una de estas dos dimensiones parece estar vinculada a diferente sexo.

El modo diferencial de manifestarse el sesgo en función del género, nos sitúa en la pista de los factores culturales determinantes de la tipificación sexual de los roles. En una cultura que todavía reserva para el hombre el rol predominante de la decisión y del manejo de la realidad, es lógico esperar que sea en estas áreas de la vida donde se manifiesten las tendencias o los posibles desajustes neuróticos de su personalidad. Y en esta misma cultura que enfatiza el "cultivo de lo emocional" para la mujer, tampoco resulta extraño encontrar que sean los ítems referidos a la expresión directa o somatizada de aspectos anímicos, los que muestren sesgo en su contra. Es lógico esperar que la inestabilidad emocional del hombre, se manifieste a través de aquellas conductas congruentes con la tipificación de su rol. Si el amplio repertorio de características masculinas incluye dominio, confianza en sí mismo, capacidad para tomar decisiones, perseguir y alcanzar éxito, agresividad...; y lo femenino, por el contrario, se caracteriza por desconfianza en sí misma, falta de capacidad de decisión y dominio,... (Martínez-Benlloch, Barberá & Pastor, 1988), cabe esperar que las tendencias neuróticas se manifiesten de un modo diferen-

cialmente específico para cada uno de los géneros. Esto explicaría los resultados de estudios que han partido de la perspectiva de género y que muestran que el contenido de los elementos que integran diversas escalas de Feminidad, más que medir supuestos rasgos de Feminidad lo que reflejan realmente es neurosis (Lunneborg, 1972; Pleck, 1975). En algunos estudios aparecen relaciones entre Feminidad y Neuroticismo (Zeldow, Clark & Daugherty, 1985) aunque los resultados no son concluyentes y hay otros estudios que no confirman estos hallazgos. Por el contrario, la alta Masculinidad aparece negativamente correlacionada con Neuroticismo (Spence & Helmreich, 1978; Spence, Helmreich & Holahan, 1979; Feather, 1985). Y en general, los principales indicadores de salud mental, entre los cuales se encuentra el Neuroticismo, dan mejor ajuste en sujetos con más Masculinidad que con más Feminidad (ver Sebastián, 1988).

La Masculinidad se ha definido tradicionalmente por su orientación *instrumental* frente a la orientación *expresiva* de la Feminidad. Los ítemes que aparecen sesgados contra los hombres parecen encajar con la orientación hacia la instrumentalidad, enfocada a la consecución de metas. Mientras que los ítemes sesgados contra las mujeres encajan mejor con la orientación *expresiva*: somatizaciones y aspectos anímicos. ¿Podría haber un *solapamiento* entre estos dos constructos psicológicos, que estaría explicando por un lado las puntuaciones más altas en Neuroticismo por parte de las mujeres, y por otro lado las correlaciones positivas del Neuroticismo con Feminidad y negativas con Masculinidad? Se necesita investigación en este sentido para poder determinar este solapamiento indirectamente apuntado en investigaciones tanto sobre Neuroticismo como sobre la perspectiva de género.

Ahora bien, una pregunta lógica ante estos resultados sería la siguiente: si aparece sesgo tanto en contra de los hombres como en contra de las mujeres, ¿Por qué decir que la escala de Neuroticismo parece estar sesgada contra las mujeres? La respuesta es simple: porque el número de ítemes sesgados contra las mujeres es doble del número de ítemes sesgados contra los hombres; esto es, la escala está más saturada de los ítemes sesgados contra las mujeres que de los ítemes sesgados

contra los hombres. Y esto con el criterio más *estrecho*, del sesgo. Si nos remitimos a la dimensión NEURO-S identificada en la escala, 8 de los 9 ítemes afectados por el componente de género resultan desventajosos para las mujeres.

Un análisis de los resultados más contextualizado culturalmente, revela una gran congruencia con el modelo dialéctico del *síndrome machista* propuesto por Martín-Baró (1987), que explica el fenómeno del *machismo*, ampliamente arraigado y debatido en América Latina (Martín-Baró, 1980; 1983). La escala N de Eysenck podría estar siendo afectada además, por un importante elemento cultural —*el machismo*—, cuyo efecto sería necesario eliminar del instrumento para poder utilizarlo en ese medio cultural (Delgado, 1994).

La hipótesis que surge de estos resultados es si el Neuroticismo no tendrá manifestaciones diferentes en hombres y mujeres, que habrá que profundizar de acuerdo a la Psicología Diferencial de los sexos, y a las manifestaciones socioculturales de género. Si bien es cierto que el Neuroticismo hace referencia a la inestabilidad emocional, su medición en la escala de Eysenck se basa en algunas de sus posibles manifestaciones: desarreglos somáticos difusos, estados de preocupación y ansiedad, sentimientos desagradables,... La cuestión es de qué manera los ítemes operacionalizan de un modo equilibrado, las manifestaciones típicas de cada género. Con los resultados obtenidos, el cuestionario que nos ocupa parece estar más cargado de las manifestaciones propias de la mujer (6 ítemes sesgados en su contra), que de las manifestaciones propias de los hombres (3 ítemes sesgados). De mantenerse estos resultados en investigaciones posteriores, cabe plantear desde esta perspectiva la necesidad de operacionalizar la dimensión de Neuroticismo de un modo diferencial para hombres y mujeres. Probablemente sería oportuno crear escalas diferenciadas para cada sexo.

CONCLUSION

Los resultados de este estudio en población salvadoreña confirman que las mujeres dan puntuaciones más altas en Neuroticismo que los hombres. Se confirma la presencia de

dos dimensiones en la escala N del EPI: NEURO-A libre de sexo, y NEURO-S ligada al sexo. Mientras en la dimensión NEURO-S las mujeres mantienen puntuaciones más altas que los hombres, en la dimensión NEURO-A no hay diferencia entre géneros.

Se ha encontrado sesgo en nueve ítems de los 24 que componen la escala N del EPI. El número de ítems sesgados en contra de las mujeres (25%) es doble del número de ítems que exhiben sesgo en contra de los hombres (12,5%). En total el 37,5% de los ítems de Neuroticismo del Eysenck Personality Inventory exhiben sesgo de género, detectado mediante la técnica TID. Las áreas de sesgo encajan con las orientaciones *instrumental* y *expresiva*, sexualmente tipificadas. Se requiere nueva investigación para confirmar estos resultados, con técnicas de detección de sesgo más precisas, y para determinar si se trata de un fenómeno ligado a la cultura salvadoreña o por el contrario, es un fenómeno tan general como la diferencia de medias entre hombres y mujeres.

REFERENCIAS

- Angoff, W.H. & Ford, S. (1973). "Item-race interaction on a test of scholastic aptitude". *Journal of Educational Measurement* 10, 95-106.
- Bem, S.L. (1981). "Gender Schema theory: A cognitive account of sex typing". *Psychological Review*, 88, 354-364.
- Bem, S.L. (1983). "Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender schematic society". *Signs* 8, 598-616.
- Berk, R.A. (Ed.) (1982). *Handbook of Methods for Detecting Test Bias*, Baltimore: Johns Hopkins University Press.
- Bickel, P.J.; Hammel, E.A.; & O'Connell, J.W. (1975): "Sex bias in graduate admissions: Data from Berkeley". *Science* 187, 398-404.
- Burril, L.E. (1982). "Comparative studies of methods". In Ronald A. Berk (Ed.). *Handbook of methods for detecting test bias*. London: The Johns Hopkins University Press (p. 161-179).
- Cole, N.S. & Moss. P.A. (1989). "Bias in test use". In Linn (Ed.). *Educational Measurement*. New York: Macmillan (p. 201-219).
- Cowles, M. Darling, M. & Skanes, A. (1992). "Some characteristics of the simulated self". *Personality and Individual Differences* 13, 501-510.
- Crocker, L. & Algina J. (1986). *Introduction to classical and modern Test Theory*. New York: Holt Rinehart and Wiston, CBS College Publishing.
- Delgado, C. (1994). "El sesgo sexual en la medición psicológica", *Revista de Psicología de El Salvador*, (en prensa).
- Dunnett, S.; Koun, S. & Barber, P.J. (1981). Social desirability in the Eysenck Personality Inventory. *British Journal of Psychology* 72, 19-26.
- Eysenck, H.J. (1947). *Dimensions of personality*. London: Routledge & Kegan Paul Ltd.
- _____. (1952). *The scientific study of personality*. London: Routledge & Kegan Paul Ltd. (Traducción española: *Estudio científico de la personalidad*, 1971. Buenos Aires: Paidós).
- _____. (1960). *The Structure of Human Personality*, London: Methuen.
- _____. (1963). The validity of questionnaires and rating assessments of extraversion and neuroticism and their factorial validity. *British Journal Psychology*. 54, 51-62.
- _____. (1967) "The biological basis of personality", Springfield, Illinois: Charles C. Thomas Publishers (Traducción

- española: *Fundamentos biológicos de la personalidad*. Barcelona: Fontanella, 1970).
- _____. (1968). *Personality Structure and Measurement*. San Diego, CA: Knapp.
- _____. (1978). *You and neurosis*. Glasgow: Collins.
- _____. (1991). "Dimensions of personality: 16, 5 or 3? Criteria for a taxonomic paradigm". *Personality and Individual Differences*. 12, 8, 773-790.
- Eysenck, H.J., Barrett, P.; Wilson, G. & Jackson, C. (1992). "Primary Trait Measurement of the 21 Components of the P-E-N-System". *European Journal of Psychological Assessment*. 8, 109-117.
- Eysenck, H.J., & Eysenck, S.B.G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.
- Eysenck, H.J., & Gudjonsson, G. (1989). *Causes and cures of criminality*. New York: Plenum Press.
- Eysenck, H.J., & Wilson, G. (1979). *The psychology of sex*. London: J.M. Dent & Sons Ltd.; (Traducción española: *Psicología del sexo*, 1981. Barcelona: Herder).
- Feather, N.T. (1985). "Masculinity, femininity, self-esteem and subclinical depression". *Sex Roles*. 12, 491-500.
- Francis, L.J. (1993). "The dual nature of the Eysenckian neuroticism scales: A question of sex differences?" *Personality and Individual Differences* 15, 43-59.
- Furham, A. (1984). "Lay conceptions of neuroticism". *Personality and Individual Differences*. 5, 95-103.
- Hambleton, R.K. (1983). "Application of item response models to criterion-referenced test item selection assessment". *Applied Psychological Measurement* 7, 33-44.
- Hambleton, R.K. (1989). "Principles and Selected Applications of Item Response Theory. In R. Linn (Ed.)". *Educational Measurement*. New York: Macmillan (p. 147-200).
- Hambleton, R.K.; Clauer, B.E.; Mazor, K.M. & Jones, R.W. (1993). "Advances in the Detection of Differentially Functioning test Items". *European Journal of Psychological Assessment* 19, 1-18.
- Hambleton, R.K. Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of items Response Theory*. Newbury Park, CA: Sage.
- Holland, P.W. & Thayer, D.T. (1988). "Differential item performance and the Mantel-Haenszel procedure". In H. Wainer & H.I. Braun (Ed.): *Test Validity*. Hillsdale, New Jersey: LEA.
- Ironson, G.H. & Subkoviak, M.J.A. (1979). "A comparison of several methods of assessing item bias". *Journal of Educational Measurement* 16, 209-225.
- Jensen, A.R. (1969). "How much can be boost IQ and scholastic achievement?" *Harvard Educational Review* 39, 1-123.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Jorm, A.F. (1987). "Sex differences in neuroticism: a quantitative synthesis of published research". *Australian and New Zealand Journal of Psychiatry* 21, 501-506.
- Kohlberg, L. (1966). "A cognitive development analysis of children's sex role concepts and attitudes". In Eleanor E. Maccoby: *The Development of Sex Differences*. Stanford, CA: Stanford University Press. (Traducción española: *Desarrollo de las diferencias sexuales*. Madrid: Marova, 1972).
- Levin, J. & Montag. I. (1987). "The effect of testing instructions for handling social desirability on the Eysenck Personality Questionnaire." *Personality and Individual Differences* 8, 163-167.

- Lodhi, P.H. & Thomas, G. (1991). "Effects of experimentally induced response sets in assessing Eysenkian dimensions of personality". *Personality and Individual Differences* 12, 811-817.
- Lord, F.M. (1980): *Applications of Item Response Theory to practical testing problems*. Hillsdale, New Jersey: LEA.
- Lord, F.M. & Novick, M.R. (1968): *Statistical Theories of Mental Test Scores*. Reading, Massachusetts: Addison-Wesley Co.
- López, F. (1984): "Adquisición del rol y de la identidad sexual." *Infancia y Aprendizaje*. 26, 25-40.
- Lunneborg, P.W. (1972): "Dimensionality of M.F.". *Journal of Clinical Psychology*. 28, 313-317.
- Maccoby, E.E. (1974): *The psychology of Sex Differences*. Stanford, CA.: Standord University Press.
- Maccoby, E.E. (1983): "Socialization in the context of the family: Parent-Child interaction". In E.M. Hetherington (Ed.): *Handbook of Child Psychology. Socialization. Personality and Social Development*. New York: Wiley.
- Martín, M.F. & Delgado, C. (1993): *Estadística Aplicada. Tratamiento informático con el programa Stat-View 512+*, Publicaciones Universidad Pontificia de Salamanca, 1993.
- Martín-Baró, I. (1980): "La imagen de la mujer en El Salvador". *Estudios Centroamericanos*. UCA, San Salvador". 380, 557-568.
- Martín-Baró, I. (1983): "Los rasgos femeninos según la cultura dominante en El Salvador". *Boletín de Psicología*. UCA, San Salvador, 3-7.
- Martín-Baró, I. (1987): "¿Es machista el salvadoreño?" *Boletín de Psicología*. UCA, San Salvador, 24, 101-122.
- Martínez-Benlloch, I.; Barberá, E. & Pastor, R. (1988); Medida de la masculinidad, feminidad y androginia psicológica. In J. Fernández (Coordinador): *Nuevas perspectivas en el desarrollo del sexo y el género*. Madrid: Pirámide (p. 125-143).
- Mischel, W. (1970): "Sex typing and socialization". In P.H. Mussen (Ed.): *Carmichael's manual of child psychology* vol.2. New York: Wiley.
- Muñiz, J. (1990): *Teoría de Respuesta a los ítems. Un nuevo enfoque en la evolución psicológica y educativa*, Madrid: Pirámide.
- Osterlind, S.J. (1983): *Test Item Bias*. Beverly Hills, CA: Sage.
- Pine, S.M. (1977): Applications of item response theory to the problem of test bias. In D.J. Weiss (Ed.): *Applications of Computerized Adaptive Testing*. Research Report 77-1. Minneapolis: University of Minnesota. Psychometric Methods Program, Department of Psychology.
- Pleck, J.H. (1975): "Masculinity-femininity. Current and alternative paradigms". *Sex Roles*. 1, 161-178.
- Poortinga, Y. & Van de Vijver, F. (1987): "Explaining Cross-Cultural Differences, Bias Analysis and Beyond". *Journal of Cross-Cultural Psychology* 3, 259-282.
- Reynolds, C. & Brown, R. (Ed.) (1984): *Perspectives on Bias in Mental Testing*. New York: Plenum Press.
- Rim, Y. (1982): "Self-monitoring, ethical position, personality, values and cognitive performance". *Personality and Individual Differences*. 3, 219-220.
- Roger, D. & Nesshoever, W. (1987): "The construction and preliminary validation of a escale for measuring emotional control". *Personality and Individual Differences* 8, 527-534.

- Rudner, L.M. & Convey, J.J. (1978): *An evaluation of select approaches for biased item identification*, Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978 (ERIC Document Reproduction Service nº ED 157 942).
- Rudner, L.M.; Getson, P.R. & Knight, D.L. (1979): *The effect of various test and item properties on five approaches to biased to biased item detection*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1979.
- Rushton, J.P.; Brainerd, C.J. & Pressley, M. (1983): "Behavioral development and construct validity: The principle of aggregation". *Psychological Bulletin*. 94, 18-38.
- Sebastián, J. (1988): "Androginia y flexibilidad de roles". En J. Fernández (Coordinador): *Nuevas perspectivas en el desarrollo del sexo y el género*. Madrid, Pirámide (p. 171-204).
- Shepard, L.A.; Camilli, G. & Averill, M. (1981): "Comparison of procedures for detecting test-item bias with both internal and external ability criteria". *Journal of Educational Statistics*. 6, 317-375.
- Shepard, L.A.; Camilli, G. & Williams, D.M. (1985): "Validity of Approximation Techniques for detecting item bias". *Journal of Educational Measurement* 2, 77-105.
- Scheuneman, J.D. (1975): A new method of assessing bias in test items. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C. April 1975 (ERIC Document Reproduction Service nº ED 106 359).
- Spence, J.T. & Helmreich, R.L. (1978): *Masculinity and Femininity, Their Psychological Dimensions, Correlates and Antecedents*. Austin: University of Texas Press.
- Spence, J.T.; Helmreich, R.L. & Holahan, C.K. (1979): "Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behavior". *Journal of Personality and Social Psychology*. 37, 1673-1682.
- TEA Ediciones (1983): EPI, *Cuestionario de personalidad*. Adaptación española. Madrid.
- Thorndike, R. (1989): *Psicometría Aplicada*. México: Limusa.
- Unger, R.K. (1979): "Toward a redefinition of sex and gender". *American Psychologist* 34, 1085-1094.
- Van de Vijver, F. & Poortinga, Y.H. (1991): "Testing Across Cultures". In R. Hambleton & J. Zaal: *Advances in Educational Testing: Theory and Applications*. Massachusetts: Kluwer Academic Publishers (p. 277-308).
- Velicer, W. & Weiner, B. (1975): "Effects of sophistication and faking sets on the Eysenck Personality Inventory" *Psychological Reports* 37, 71-73.
- Zeldow, P.B. & Clark, D. & Daugherty, S.R. (1985): "Masculinity, femininity, type A behaviors, and psychological adjustment in medical students". *Journal of Personality and Social Psychology* 48, 481-492.

Carmen Delgado Alvarez
 Facultad de Psicología
 Universidad Pontificia de Salamanca
 C/Compañía, 1/37008 Salamanca, España
 Fax: (23) 262456
 E-mail: cdelgado@gugu.usal.es