

## Differential Item Functioning

P.W. Holland y H. Wainer, Eds.

Hillsdale, NJ: Lawrence Erlbaum Associates,  
1993, 453 págs.

1.—Los que están interesados en el tema lo conocen. 2.—Los que no lo conocen, no están interesados en el tema. Si la primera proposición es cierta y alguien lee esta revisión, debo suponer que busca más el juicio que la información. Y el buen juicio se encuentra en el buen juez. Quien busque jueces buenos y buenos juicios, los encontrará en el *Journal of Educational Measurement* (Clauser y Hambleton, 1994) y en *Applied Psychological Measurement* (Millsap, 1993). Pero, si está interesado en el tema, ya los conocerá. ¿Para quién se escribe? ¿Para el fuego, los hongos y el polvo de las bibliotecas? En el límite, con certeza. Tenemos que suponer, por lo tanto, que la segunda proposición es falsa, y que existe alguien que está interesado en el tema, que no conoce ni el libro ni las revisiones citadas, y que, además, lee *Psicothema*. A este lector extraviado va dedicada la revisión del libro. Al conjunto vacío.

Es muy difícil leer *Differential Item Functioning* y no pensar en Wittgenstein. Es muy difícil, pensando en Wittgenstein, no mostrar parte del contenido del libro a través de un análisis del lenguaje. ¿Cuál es el objetivo de la filosofía? Según Wittgenstein, mostrar a la mosca el orificio de salida de la botella. Dicho de otro modo, desenredar el malentendido y confusión de los usos lingüísticos que están en la base de los problemas filosóficos, disolviéndolos de esta forma. Sin embargo, el proceso de clarificación conceptual, aunque objetivo primario de la filosofía analítica, está indisolublemente ligado a cualquier ámbito de la actividad científica. El lenguaje, en ciencia, no es baladí: utilizar

los mismos términos con diferentes significados lleva, como mínimo, a confusión. La historia de la investigación sobre el sesgo es, a este respecto, ejemplificadora. Gran parte de la controversia sobre el particular, recuérdese el libro de Jensen (1980), fue debida a la utilización de la misma palabra, sesgo (*bias*), con diferentes significados. Una cosa es el significado y las connotaciones sociales de la palabra sesgo, y otra su significado estadístico. No obstante en las discusiones, sobre todo por parte de los legos, se equiparaba con frecuencia el término técnico sesgo al concepto ético de injusticia (*unfair*). A todo ello bien mezclado, por si no fuera poco, le solía acompañar la falacia igualitarista. La argumentación final solía adoptar esta forma: Supuesto trascendental: todos los hombres son iguales. Comprobación empírica: no todos los grupos humanos tienen el mismo desempeño promedio en los tests. Conclusión: los tests están sesgados contra algunos grupos humanos; por tanto, si los tests hacen a los hombres, que son iguales, desiguales, los test son injustos, ya que la justicia está con la igualdad de oportunidades. Contra esta amalgama de «*les droits du citoyen*» con el «*american way of life*», contra la confusión de planos y, en general, contra la estupidez manifiesta, se enfrentó Jensen. Hasta aquí una parte de la historia, la peor, vamos con la otra. Si la confusión, como dijimos, surge de la utilización indiferenciada de los distintos usos lingüísticos del mismo término, el proceso de clarificación conceptual empieza con la creación de nuevas expresiones para cada uso lingüístico. Así el término «sesgo de los items» fue sustituido por el más largo pero preciso «funcionamiento diferencial de los items» (DIF). Decir que un item presenta DIF, es decir, simplemente, que muestra diferentes propiedades estadísticas en diferentes grupos. Si es el test en su conjunto el que presenta propiedades es-

tadísticas distintas en cada grupo, se habla de «funcionamiento diferencial del test» (DTF). La palabra sesgo, tanto de los items como del test, se reserva ahora a las situaciones en que se puede establecer relación entre el funcionamiento diferencial y el constructo que se pretende medir. Dicho brevemente, sólo se puede hablar de sesgo en términos de validez de constructo. Decir que un item/test está sesgado implica necesariamente un funcionamiento diferencial entre grupos, pero indica además que no mide lo que pretende medir o que mide más cosas de las que pretende medir. Pero no de toda diferencia entre grupos se sigue la existencia de sesgo. Si todos los items del test midiesen sólo la habilidad pretendida, cualquier diferencia entre grupos reflejaría sólo el impacto (*impact*) de ese item/test, no el sesgo. Como señala Akerman (1992), sólo si dos grupos tienen diferentes distribuciones en las habilidades que no se pretenden medir (*nuisance abilities*), y los items del test son capaces de medir esas múltiples dimensiones, y la estimación de esas habilidades se concreta en una única medida (puntuación en el test), se dan las condiciones necesarias para que se produzca sesgo. Y es así, a través de la elucidación de los distintos usos lingüísticos que se han dado a la palabra sesgo, y a su cristalización en nuevas expresiones (DIF, DTF, impacto), como hemos llegado a la raíz del concepto: sólo se puede explicar y hablar de sesgo desde una perspectiva multidimensional. Quien esté interesado en los efectos tendrá que hablar de DIF o DTF, y lo podrá hacer con un enfoque meramente estadístico. Quien esté interesado en las causas hablará de sesgo, y lo tendrá que hacer desde una perspectiva multidimensional, a la vez que deberá llenar sus palabras de contenidos y procesos psicológicos.

Se preguntarán, ¿y esta digresión a cuenta de qué? ¿acaso ha entrado la filo-

sofía analítica a saco en el campo psicométrico? No, pierdan cuidado, el libro en cuestión es un libro estrictamente psicométrico. Editado en 1993, tuvo su origen en las actas de un congreso sobre el DIF celebrado en 1989, y copatrocinado por el *Educational Testing Service* y el *Air Force Human Resources Laboratory*. El volumen se divide en cuatro secciones: «Introducción y antecedentes», «Metodología estadística», «Aspectos aplicados e investigaciones empíricas» y «Otros aspectos».

### Introducción y antecedentes

La parte introductoria se abre con un capítulo de Angoff sobre la historia y evolución de los principales métodos surgidos para evaluar el DIF, desde el método delta, hasta los métodos desarrollados en la TRI, pasando por los procedimientos Jic cuadrado, el Mantel-Haenszel y la estandarización. No se trata obviamente de una descripción extensa de los mismos, sino de un análisis de las ventajas que cada método iba procurando sobre sus antecesores, y de las características comunes e inconvenientes de unos y otros. Con la misma claridad y acierto son tratados aspectos conceptuales relacionados, como la circularidad inherente al utilizar en los análisis las puntuaciones del test como estimación de la habilidad de los sujetos, el supuesto de unidimensionalidad del test que subyace a todos los procedimientos, o las diferencias entre sesgo y DIF, entre otros. Es un excelente capítulo introductorio que debería ir seguido del escrito por Camilli, otro veterano del sesgo, bajo el breve título de «The case against item bias detection techniques based on internal criteria: Do item bias procedures obscure test fairness issues?». En él se hace un análisis de la evolución conceptual del sesgo del test y de los items, a través de la presentación y análisis de los puntos de vista que los más notables investigadores sobre el tema

han tenido a lo largo del tiempo. Incomprendiblemente, este capítulo cierra el libro.

### Metodología estadística

Si alguien no leyera el prefacio del libro, y llegando hasta aquí tuviera dudas sobre quién o qué organización estaba detrás del mismo, se le despejarían inmediatamente. El grueso de los capítulos están dedicados a dos procedimientos: el Mantel-Haenszel y la estandarización. De ellos se pueden encontrar descripciones y comparaciones detalladas (Dorans y Holland), estudios de simulación para ver como se ven afectados por diversos factores en la detección del DIF (Donoghue, Holland y Thayer), y un procedimiento para valorar la estabilidad del índice de DIF que se deriva del procedimiento Mantel-Haenszel (Longford, Holland y Thayer). Los métodos derivados desde el marco de la TRI son tratados también con cierta extensión. Así por ejemplo, Thissen, Steinberg y Wainer, describen cuatro métodos para detectar el DIF, además de un modelo para detectar el «funcionamiento diferencial de las alternativas», en la línea del análisis del «funcionamiento diferencial de los distractores» propuesto por Green, Crone y Folk (1989). Dentro del marco de los modelos de rasgo latente, Shealy y Stout presentan un modelo multidimensional no paramétrico para explicar el sesgo y el funcionamiento diferencial de los test. Es a mi juicio el único intento extenso y riguroso de explicar el funcionamiento diferencial, tanto del test como de los ítems. La explicación del DIF es la misma que ya en su día expusiera Akerman (1992), el DIF se produce cuando hay ítems multidimensionales en un test que se pretende unidimensional y existen diferentes distribuciones entre grupos en alguna de las habilidades que no se pretenden medir. Ellos le

añaden una teoría formal y un intento de conectar el comportamiento diferencial de los ítems con el del test. Además, a posteriori, teniendo como fundamento el modelo formal aquí presentado, han desarrollado un procedimiento para evaluar el DTF y el DIF, el SIBTEST (Shealy y Stout, 1993), que será por sus características (procedimiento basado en un modelo multidimensional no paramétrico, que proporciona una estimación de la cantidad de DTF/DIT presente y un test de significación, que permite considerar el comportamiento diferencial del test y de los ítems conjuntamente, y que no es costoso computacionalmente) uno de los procedimientos que más investigaciones e interés generarán en el futuro. Lo afirmo con la seguridad de quien apuesta a caballo ganador. Sin embargo, el SIBTEST tiene una limitación que comparte con el Mantel-Haenszel y la estandarización: es incapaz de detectar el sesgo no uniforme. Aunque para decir toda la verdad, recientemente Mazor, Clauser y Hambleton (1994) han propuesto una modificación del procedimiento Mantel-Haenszel que parece ser eficaz en la detección de este tipo de sesgo. Otro de los procedimientos que sí pueden detectar el sesgo no uniforme es la regresión logística (Swaminathan y Rogers, 1990). En el libro, lamentablemente, tan sólo se cita.

### Aspectos aplicados e investigaciones empíricas

Alguien podría pensar que no existe ninguna relación entre el psicoanálisis y las investigaciones sobre el DIF, y se equivocaría. Como sardónicamente señala Lloyd Bond, en ambas áreas se puede explicar todo, pero no se puede predecir nada. Un ejemplo de este proceder lo tenemos en el capítulo escrito por O'Neill y McPeck en el que mediante análisis *post hoc* se señalan algunas características de

los items y de los test que parecen estar relacionadas con el funcionamiento diferencial de los mismos. Más allá de este enfoque a posteriori y meramente correlacional, Schmitt, Holland y Dorans proponen el uso de una metodología experimental para testar hipótesis sobre las posibles causas del DIF. Si el primer enfoque es exploratorio y correlacional, este enfoque es confirmatorio y causal. Los restantes capítulos tienen, tanto por la forma como por el contenido, un alto valor práctico. Están constituidos por preguntas de indudable interés aplicado seguidas de breves respuestas, constituyéndose en una especie de guía de «todo lo que usted siempre quiso saber acerca del DIF y nunca se atrevió a preguntarlo». Como muestrario, he aquí algunas: ¿Cómo afecta la eliminación de los items con DIF al promedio de las puntuaciones en el test? ¿Cuál es el tamaño de muestra requerido para realizar análisis sobre el DIF? ¿Qué grupos deben ser analizados y cómo definirlos?.

#### Otros aspectos

Pocos abogados coincidirían con los editores en incluir en este apartado un capítulo que describe el marco legislativo norteamericano en el que tienen que desenvolverse las investigaciones sobre DIF (McAllister). Parece ser que no se les pidió consejo. Los otros capítulos que integran esta sección son el de Camilli, comentado anteriormente, y el escrito por Ramsay sobre la historia y experiencia que tiene el Educational Testing Service en el proceso de revisión de los items para eliminar cualquier aspecto de los mismos que pueda ser considerado ofensivo por las minorías, las mujeres y los individuos con discapacidades, o inducirles, por motivos independientes de la habilidad evaluada, a un peor desempeño en el test. Aunque los temas trata-

dos aquí, excepto el de Camilli, están bien calificados como accesorios o auxiliares, no dejan por ello de tener el interés de mostrarnos el objetivo y la finalidad social última de las investigaciones sobre el comportamiento diferencial de los test y de los items.

Como conclusión, señalar que el libro que nos ocupa se constituirá, como aconteció en su década con el de Berk (1982), en la guía básica de las investigaciones sobre el sesgo de los noventa, a la vez que sirve como prospección de futuro que permite entrever por donde ocurrirá la investigación en lo que nos queda de siglo. Y es que, después de todo, un libro en el que se cita a Aristóteles, a Spinoza y a J.S. Mill, difícilmente puede ser un mal libro. No me queda más que decir. Confío que sabrán perdonarme que, a fuerza de andar por estos parajes y sobre estos temas, no sea del todo imparcial y mi mano escorde, arrastrada por mi pensamiento, a escribir sobre aquello que más me gusta. Y si juzgan el espacio dedicado a cada cosa, sabrán que es.

#### Referencias

- Ackerman, T.A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29 (1), 67-91.
- Berk, R.A. (1982). *Handbook of methods for detecting test bias*. Baltimore: The Johns Hopkins University Press.
- Clauser, B.E. y Hambleton, R.K. (1994). Recensión de *Differential Item Functioning* editado por P.W. Holland y H. Wainer. *Journal of Educational Measurement*, 31(1), 88-92.
- Green, B.F., Crone, C.R. y Folk, V.G. (1989). A method for studying differential distractor functioning. *Journal of Educational Measurement*, 26(2), 147-160.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Mazor, K.M.; Clauser, B.E. y Hambleton, R.K. (1994). Identification of nonuniform

differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement*, 54(2), 284-291.

Millsap, R.E. (1993). Recensión de *Differential Item Functioning* editado por P.W. Holland y H.Wainer. *Applied Psychological Measurement*, 17 (1), 101-103.

Shealy, R. y Stout, W. (1993). A model-based standarization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58, 159-194.

Swaminathan, H. y Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.

Revisado por  
**Angel M. Fidalgo Aliste**  
Facultad de psicología.  
Universidad de Oviedo.

## **Enuresis infantil. Un problema con solución.**

Bragado Alvarez, M. C. (1994)  
Madrid, Eudema. 112 pp.

Cuando me disponía a iniciar el comentario de este libro de la profesora Carmen Bragado acudieron a mi memoria las palabras con las que hace unos pocos meses comunicaba su aparición a mis alumnos de la especialidad de psicología clínica: «Hoy es uno de esos días en los que por fuerza uno ha de ponerse contento, pues acaban de entregarme un buen libro. Digo que es un buen libro porque está muy documentado, es riguroso, está bien elaborado, es breve, aborda un tema importante en la clínica y da orientaciones precisas para la intervención. Su título es *Enuresis infantil. Un problema con solución*. Y su autora es la profesora Bragado.

En efecto, estamos ante un libro muy documentado porque su autora ofrece una información exhaustiva sobre la enuresis, a partir de una bibliografía relevante, amplia y actualizada, tanto sobre su conceptualización y su incidencia como sobre el proceso y problemas en el control de la micción, o sobre los diferentes sistemas de tratamiento y su eficacia.

Es también un trabajo riguroso pues aborda minuciosamente cada aspecto de la enuresis, presentando el estado de la cuestión a través de una exposición crítica del conocimiento actual biológico y psicológico acerca del proceso de control de la micción, apoyándose en todo momento en hechos, en datos contrastados, obtenidos de numerosas investigaciones ya sean epidemiológicas, sobre el funcionamiento fisiológico relacionado con la micción, o investigaciones de las diversas modalidades de intervención. De alguna manera debía reflejarse en el libro la labor rigurosa de investigación que Carmen Bragado viene realizando sobre la enuresis a lo largo ya de dos décadas.

Asimismo, es una obra bien elaborada, acorde con las dotes docentes de su autora y

fruto de la experiencia en la brega diaria con alumnos, pues va presentando la información de lo general a lo particular y de la fundamentación teórica a la aplicación práctica, desmenuzando cada uno de los aspectos conceptuales o prácticos que un lector formado en psicología puede necesitar para comprender el fenómeno de la enuresis y su control.

Así, se distribuye el libro en cinco capítulos, a los que se les añade unos apéndices. En el primer capítulo se expone qué puede considerarse enuresis, su definición tradicional, la matización-complicación que implica el criterio del DSM III-R al incluir la micción «intencional», los diversos tipos que se han establecido, y su incidencia en la población actual española a la luz de diversas investigaciones epidemiológicas recientes.

El capítulo segundo aborda los mecanismos fisiológicos de la micción, el proceso evolutivo que se sigue en el control de los esfínteres y una exposición crítica de las diversas causas que se han aducido para explicar el fenómeno de la enuresis nocturna, de entre las que la autora asume la hipótesis del déficit de aprendizaje.

Los capítulos tercero y cuarto se dedican a presentar los tres tratamientos psicológicos objeto de mayor investigación: el Método de Alarma de Mowrer y Mowrer, el Entrenamiento en Retención de Kimmel y Kimmel, y el Entrenamiento en Cama Seca de Azrin, Sneed y Foxx. Respecto a cada uno de ellos, Bragado va describiendo su fundamentación teórica, el procedimiento para poder llevarlo a cabo, los datos correspondientes a su eficacia, y algunas indicaciones para solventar los problemas que pueda plantear su aplicación.

El último capítulo está destinado a exponer un modelo de intervención psicológica en los problemas de enuresis. En él la autora nos describe paso a paso toda la secuencia de la actuación que ha de desarrollar el psicólogo desde el primer contacto con los padres hasta que finaliza el seguimiento: el proceso de evaluación y los procedimientos más adecuados para llevarla a cabo, la planificación del tratamiento

(el procedimiento de elección, cuándo y cómo iniciarlo y finalizarlo), y el modo de realizar el seguimiento. Tanto en este capítulo como en los dos anteriores, el modo de plantear la actuación práctica revela la amplia experiencia que la autora posee al enfrentar los problemas de enuresis en la clínica del día a día.

Finalmente, se completa toda la información anterior con unos apéndices muy útiles, pues contienen las instrucciones que se han de dar por escrito a los padres para llevar a cabo el Entrenamiento en Cama Seca (apéndice 1) y para la utilización del aparato de alarma del tipo de «alfombrilla» (apéndice 2), y diversos instrumentos de ayuda en la evaluación: una guía de entrevista infantil, la Escala de tolerancia materna ante la enuresis de Morgan y Young, y hojas de registro para el curso del tratamiento y para el seguimiento (apéndice 3).

Después de todo lo expuesto, el lector quizá se pregunte con perplejidad cómo puede caber todo lo mencionado anteriormente en un libro de apenas 112 páginas y de un formato «de bolsillo». La respuesta ha de encontrarse en su apariencia. En efecto, estamos ante un libro pequeño en su apariencia y amplio en su contenido. No obstante, esta aparente contradicción se debe a que la edición realizada no puede considerarse precisamente que haya sido muy afortunada (tipo de letra muy pequeña y sin espacio en blanco entre los párrafos) pues comprime en exceso un texto denso en información y ya de por sí concentrado, merecedor de una edición más acorde con su contenido.

En resumen, este libro no sólo dice mucho y bien sobre la enuresis, demostrando que efectivamente es un problema con solución desde un proceder psicológico científico como es la terapia de conducta, sino que también refleja perfectamente a su autora, pues deja traslucir su dedicación y esfuerzo cotidiano a lo largo de muchos años en la triple vocación, inseparable en ella, docente, clínica e investigadora.

Revisado por: Luis María Llavona.  
Universidad Complutense.