

El análisis del DIF en escalas de características profesionales: una comparación entre distintos métodos¹

Ana Paula Porto Noronha
Acácia Aparecida Angeli dos Santos
Fermino Fernandes Sisto

Universidade São Francisco, campus Itatiba-SP, Brasil

Recibido: 1 de febrero del 2012 / Aprobado: 27 de junio del 2013

El objetivo de este artículo es comparar tres métodos para verificar el funcionamiento diferencial de los ítems (DIF) en relación con la variable sexo. Participaron 1247 estudiantes (59,1 % del sexo femenino), con edades entre 14 y 73 años ($M=21,76$; $DP=7,07$), que respondieron colectivamente a la Escala de Característica y Ambientes Profesionales, en una única sesión y en aula. Los métodos mostraron diferentes resultados y la correlación más alta entre ellos fue de 0.74 y la más baja de 0.42. Así, los resultados indicaron discrepancias, dejando claro que algunos métodos son más flexibles y otros más conservadores. Sin embargo, ellos también sugirieron la necesidad e importancia de analizar el DIF para ampliar las informaciones sobre los aspectos psicométricos de las escalas y otros test de evaluación psicológica.

orientación profesional / test psicológicos / diferencias entre sexos

DIF analysis on a scale of professional characteristics: a comparison of different procedures

This study is aimed at comparing three methods to verify whether the items had differential functioning regarding the variable of sex. 1247 students (59,1 % women), aged between 14 and 73 years ($M=21,76$, $DP=7,073$), responded the Escala de Característica y Ambientes Profesionales. They were collectively tested in their classrooms in one session. The methods resulted in different outcomes and the correlations varied from .74 (highest) to .42 (lowest). The outcomes indicated discrepancies, showing that some methods are more flexible and other more conservative. However, they also suggest the importance and need of analyzing the DIF to enlarge the information about the psychometric characteristics of the tests and other examinations used for psychological assessment.

professional guidance / psychological test / sex differences

1 Traducción del portugués por Ana Bendezú.
Correo electrónico: Ana.noronha@usf.edu.br

El trabajo ha sido comprendido como una etapa fundamental del desarrollo humano, y su definición ha sufrido variaciones desde la revolución industrial del siglo XIX. En esa época la actividad del trabajador se resumía en la confección de artefactos manuales únicos; pero posteriormente se dieron grandes transformaciones en las relaciones hombre-trabajo en la industrialización (Pérez & Jaramillo, 2007). Es importante recordar, también, el largo periodo en que no estaban previstas las elecciones profesionales variadas, en oposición a los tiempos modernos en los que se encuentra una amplia red de opciones.

Escoger la actividad profesional, entretanto, implica entrar en contacto con aspectos relevantes, dentro de ellos, los campos de interés, el descubrimiento del significado del trabajo, la importancia de la familia en el proceso, las condiciones socioeconómicas, las características de quien escoge, además del mercado de trabajo, siempre en constante transformación. El proceso de Orientación Profesional (OP), tal como es realizado en el Brasil, considera los elementos citados con la finalidad de promover resultados más eficaces (Melo-Silva, Noce & Andrade, 2003).

Como instrumentos auxiliares del proceso de recolección de informaciones, en la OP están las pruebas psicológicas, cuyas premisas, de acuerdo con Anastasi y Urbina (2000), versan sobre la identificación de diferencias indi-

viduales. Las autoras defienden que el uso de los instrumentos de evaluación profesional permite investigar cuan estrechamente relacionados están los intereses de algunos individuos con los de las personas que tienen éxito en determinadas ocupaciones, así como relacionar los intereses con otros constructos psicológicos.

En el Brasil, tal como es destacado por Teixeira, Castro y Cavalheiro (2008) en el Sistema de Avaliação dos Testes Psicológicos (Satepsi) (Conselho Federal de Psicologia, 2009), hay pocos instrumentos disponibles para su uso en este contexto. Ese hecho, por sí solo, justifica el desarrollo de investigaciones sobre el tema.

Con este deseo se ofrece el presente estudio, que versa sobre el análisis de las diferencias entre hombres y mujeres en un instrumento que se destina a evaluar las preferencias por características y ambientes de trabajo. Considérase un test con evidencia de validez cuando ofrece medidas idénticas para personas con niveles semejantes de la habilidad que el test se propone medir. No obstante, no siempre la puntuación obtenida se refiere solo al nivel de habilidad de las personas en la variable investigada, sino también a características irrelevantes, tales como pertenecer a diferentes grupos culturales, étnicos, entre otros; en el caso específico de este estudio, ser de sexo masculino o femenino. Ese tipo de interferencia constituye una evidente y definida causa para inviabilizar

el instrumento, pues las puntuaciones sufren interferencias significativas de variables que no son la habilidad que se quiere medir, dificultando su interpretación (Sisto, 2006).

A fin de presentar un breve panorama de las investigaciones ya desarrolladas sobre el tema, serán presentados algunos trabajos considerados más pertinentes en el sentido de informar acerca de la importancia de la variable 'sexo' en la elección de actividades profesionales. Para facilitar la configuración del panorama se optó por organizar las investigaciones describiendo primero las extranjeras y después las nacionales, tal como se presenta en los párrafos que siguen.

Lent, López y Bieschke (1991) investigaron si había diferencia de acuerdo con el sexo en cuanto a los intereses por la matemática, considerando que en la época había pocas investigaciones que abordaban el tema. Los autores concluyeron que los intereses por dicha materia variaban en función del sexo, de tal modo que los hombres presentan medias más altas. Para los investigadores, las diferencias se deben al hecho de una mayor exposición a los temas afines, lo que puede crear más oportunidades de desarrollo con la matemática.

Tomando como referencia el modelo ofrecido por el Self Directed Search (SDS), Holland, Johnston y Asama (1994) investigaron a 298 adultos que buscaban un *workshop* sobre carre-

ras. Los autores analizaron la asociación entre intereses y personalidad, de acuerdo con las perspectivas teóricas del Riasec, o sea, del modelo hexagonal tipológico propuesto por el autor, en el cual figuran seis tipos: Realista, Investigativo, Artístico, Social, Emprendedor y Convencional; y el de los Cinco Grandes Factores. En lo que se refiere a los intereses, focalizados en el presente estudio, los resultados revelaron diferencias importantes por sexo en casi todos los tipos de Riasec, con excepción de Emprendedor, la más significativa de las cuales fue la relativa al tipo Realista, con medias más elevadas para los hombres.

La diferencia de sexos relacionada con la elección por ciencias fue observada por Collier, Spokane y Bazler (1998). Por tanto, fueron investigados 1501 alumnos de cuarto y décimo años de los Estados Unidos (809 mujeres y 692 hombres). Los resultados evidenciaron la existencia de diferencias significativas entre los intereses por las carreras de ciencias, revelando que las niñas tenían menor preferencia por ellas, con excepción de las carreras que se asociaban a la investigación de algún componente social.

También Cejka e Eagly (1999) enfatizaron que para alcanzar el éxito en algunas áreas, es importante que la persona tenga determinadas características. Así, los autores recurrieron a las diferencias que han sido referidas a perso-

nas del sexo masculino, cuando fueron comparadas con las del sexo femenino, consideradas como prevalentes en uno y otro. En las mujeres, normalmente se destaca la imaginación, la intuición, el aspecto artístico, la percepción, la creatividad, la mayor fluencia verbal y mayor facilidad de contacto, entre otras. Y los hombres son más analíticos, buenos con números, cálculos, raciocinio, abstracción y resolución de problemas. Afirmaron que en algunas ocupaciones la mayor predominancia es la del sexo femenino, como dentistas y secretarias, y otras son predominantemente masculinas, como es el caso del gestor, administrador, mesero y farmacéutico.

Con vistas a evaluar la relación entre constructos investigados cuando se trata de los procesos de asesoramiento de carrera, entre ellos los intereses profesionales, Rottinghaus, Lindley, Green y Borgen (2002) realizaron una investigación con 365 estudiantes. Para medir los intereses fue utilizado el Strong Interest Inventory, y los resultados mostraron diferencias significativas entre los sexos en lo que se refiere a los tipos Realista, Investigativo, Artístico y Social. En los dos primeros, los promedios mayores fueron de los hombres y en los dos últimos los de las mujeres.

En Brasil, recientemente han sido publicadas diversas investigaciones que abordan esa cuestión. Así, entre los estudios brasileños recuperados destaca el de Lemos (2001), que comparó veinte niñas y veinte niños en cuanto al Pro-

cedimiento de Diseños de Profesionales con Historias. Los niños revelaron su preferencia por el modelo solitario de trabajo, con gran exigencia en la dedicación y calificación profesionales. Para ellos, las ganancias materiales y el estatus que la profesión confiere son elementos más relevantes que las actividades profesionales en sí. Las niñas, a su vez, se preocuparon más por la realización profesional, siendo el retorno financiero la consecuencia. Ellas prefieren las actividades profesionales que permitan el trabajo en equipo, la contribución para la mejora de la sociedad mediante el ejercicio profesional, y, finalmente, aquellas en que puedan contribuir a las relaciones sociales.

La comparación de 493 estudiantes universitarios, evaluando si había una diferencia relacionada con la variable sexo en la manera como ellos percibían sus carreras profesionales, fue realizada por Lemos y colaboradores (2005). Por tanto, los autores utilizaron un cuestionario para evaluar inclinaciones profesionales, identificando que los hombres destacaban en el Emprendimiento y Autonomía y libertad. La inclinación profesional para ser un emprendedor revela individuos motivados para superar desafíos y mostrar sus competencias y habilidades, mientras la Autonomía y libertad implica alta inversión en la carrera autónoma. Las mujeres presentaron mayor inclinación por el Altruismo y la Estabilidad y calidad de vida. En el

altruismo es posible identificar la tendencia en cuidar del otro y la busca del entendimiento mutuo. Al lado de eso, la Estabilidad y calidad de vida sugiere la preocupación en atender las necesidades de la vida profesional y de la vida familiar conjuntamente.

Sartori (2007) evaluó las preferencias profesionales por medio de la Escala de Asesoramiento Profesional (EAP) y del SDS, de 132 estudiantes de enseñanza media de entre 14 y 19 años. Hubo correlaciones significativas con magnitud de débil a moderada, entre las dimensiones del EAP y los tipos do SDS. Los resultados también indicaron la diferenciación por serie y, en relación con la diferencia entre sexos solamente en Ciencias Biológicas y de la Salud hubo diferenciación significativa, con mayor media para el sexo masculino, evidenciando que los hombres tuvieron más preferencia por el área que las mujeres.

Haciendo uso de un instrumento que evalúa las inclinaciones profesionales, el Test de Fotos de Profesiones (BBT) (Welter 2007) comparó el desempeño de 204 jóvenes y adultos que buscaban orientación profesional o reorientación de carrera. Los resultados encontrados indicaron inclinaciones femeninas y masculinas distintas. Las mujeres predominantemente escogieron positivamente los factores Z, Z', S', O y W, que sucintamente tratan de la 'preocupación con la estética', del 'sentido social', de la 'sociabilidad', de la 'necesidad de

alimento' y de la 'feminidad'. Los hombres prefieren los factores S', S, V', Z' y G. Especialmente, los tres últimos representan 'razón y lógica', 'necesidad de mostrar' e 'intuición e imaginación'.

Nunes y Noronha (2009) estudiaron las asociaciones entre intereses y trazos de personalidad en 115 jóvenes que hacían parte de un programa de orientación profesional. Las diferencias encontradas en lo que se refiere a los intereses, por medio de la aplicación del SDS, revelaron que las mujeres tienen más intereses por el tipo Artístico, mientras que los hombres por el Realista.

Con vistas a analizar las diferencias de promedios entre los instrumentos EAP y SDS, en lo que respecta al sexo, Sartori, Noronha y Nunes (2009) investigaron a 177 estudiantes de la enseñanza secundaria. En ambos instrumentos la diferencia fue observada, siendo que en la EAP los hombres tuvieron promedios mayores en Ciencias Exactas y las mujeres en Ciencias Biológicas y de la Salud. Ya en relación con el SDS, los hombres tuvieron mayor preferencia por los tipos Realista y Emprendedor, al paso que las mujeres por lo Social.

Las relaciones entre la EAP y la Escala de Características de Actividades Profesionales (ECAP) fueron estudiadas por Santos, Noronha y Sisto (2008a). Entre los principales resultados obtenidos a partir de la aplicación en 122 jóvenes con promedio de edad en 15,79 años, se destaca una correla-

ción entre la dimensión Artes y Comunicación con el ítem ‘desarrolla trabajo de manera creativa’ ($r=0.46$), de la dimensión Ciencias Agrarias y Ambientales con ‘trabajar junto a la naturaleza’ ($r=0.55$), de las Actividades Burocráticas con ‘ejecutar proyectos’ ($r=0.32$) y de las Ciencias Humanas y Sociales Aplicadas con ‘trabajar usando su capacidad de expresión verbal’ ($r=0.38$). Aunque el estudio haya tenido un carácter exploratorio, permitió comprender las preferencias y las asociaciones de aquellos jóvenes al momento de la opción profesional.

La selección de publicaciones relacionada indica que en los estudios sobre profesiones son bastante comunes los resultados con evidencias de que el sexo es una variable que diferencia a las personas. En ese contexto, se puede preguntar si esa variable diferencia a las personas en cuanto a la opción profesional o si los ítems del instrumento direccionan los resultados por facilitar más respuestas de personas de un sexo en detrimento de otro. En otras palabras, se cuestiona si ya hay una parcialidad en los ítems para que ese resultado (diferencia entre los sexos) ocurra o si no la hay, lo que podría ser atribuido a diferencias realmente producidas en función del sexo.

En ese contexto, se harán análisis para detectar los ítems con probabilidad de aciertos diferentes entre subgrupos

de una población, cuyas personas poseen un nivel similar de habilidad en la variable medida. Los estudios con vistas a detectar o neutralizar ese efecto, hoy denominados de *funcionamiento diferencial de los ítems* (DIF), son de fundamental importancia por sus implicaciones éticas, sociales y jurídicas, pues la utilización de test sin ese tipo de estudio puede subestimar las capacidades de ciertos grupos en función de características diferenciadoras irrelevantes para la habilidad en cuestión (Camilli, 1993; Camilli & Shepard, 1994; Holland & Thayer, 1988).

En este estudio se dio preferencia a la investigación con base en un criterio interno, que se refiere a las propiedades sicométricas de los ítems. Así, se buscó responder si los ítems del instrumento poseen el mismo comportamiento estadístico (o equivalencia de medida) cuando se comparan subgrupos de sujetos pertenecientes a la misma población (Sisto, 2006). Si hubiera equivalencia de medida se interpretará que no hay funcionamiento diferencial de los ítems (DIF).

Ese tipo de análisis será hecho usando tres modelos diferentes, con vistas a una comparación de los resultados. Uno de ellos derivado de la Teoría Clásica de los Tests (TCT), otro del modelo de Mantel-Haentzel y, por fin, el modelo de Rasch (vehiculado por el Winsteps).

MÉTODO

Participantes

Participaron 1247 individuos, con edades entre 14 y 73 años (media=21,76; DP=7,073), de los cuales 59,1 % de sexo femenino y 40,9 % de sexo masculino. Con relación a la escolaridad, 72,2 % cursaba educación superior y 27,8 % educación media. Todos los participantes provenían de instituciones particulares.

Instrumento

Se utilizó una ECAP, cuyos estudios están siendo ejecutados (Santos, Noronha & Sisto, 2008b). La versión inicial está compuesta por 35 características de ambientes profesionales con formato Likert, y las respuestas deben variar de mucho (5) a no tiene (1), según cuánto representa sus preferencias cada uno de los ítems.

La escala fue construida a partir de la relación de 65 profesiones descritas en guías ocupacionales y descripciones de perfiles profesionales de varias instituciones de enseñanza superior, divulgados en internet. Una vez reunidas las informaciones se llegó a una relación compuesta por 35 ítems que describen peculiaridades de las actividades profesionales.

Se desarrollaron estudios preliminares para investigar las posibilidades de factorialidad, encontrándose un KMO de 0.79 en el test de esfericidad de Bar-

lett, significativo ($\chi^2=6935,21; p<0.001$). Una rotación Promax, por medio de inspección visual al *scree plot* permitió identificar tres factores, que traducían las características físicas del ambiente de trabajo; una de las características de las actividades profesionales es la relacionada con la rutina. El alfa de Cronbach para la escala general reveló un coeficiente de 0.73.

Procedimientos

Luego de la aprobación del proyecto por el Comité de Ética, se realizaron los contactos con las instituciones. Se combinaron las respectivas direcciones, los mejores horarios y los días de la semana para la aplicación. Para los estudiantes de menor edad, los Términos de Consentimiento Libre y Esclarecido (TCLE) fueron enviados a los padres, a fin de obtener la autorización para el estudio. Los mayores de edad que estuvieron de acuerdo en participar firmaron los TCLE. Las aplicaciones ocurrieron en las instituciones de enseñanza visitadas de forma colectiva. El tiempo de realización no pasó de 15 minutos.

Procedimientos de análisis de los datos

En el modelo relacionado con la Teoría Clásica de los Tests (TCT), la puntuación bruta ha sido considerada aceptable para informar sobre la habilidad, así como la estadística suficiente para detectar dificultades del ítem y el número

de quienes respondieran correctamente ese ítem. Para controlar la habilidad de las personas en el instrumento se separaron cuatro grupos por niveles de habilidad, usando la técnica del cuartil; en todos los grupos se estudió el nivel de significancia de las diferencias en razón del sexo de las personas, ítem por ítem.

Entre los modelos estadísticos más usados para detectar ítems con el DIF está el de Mantel-Haentzel. Además de poseer pocas dificultades de cálculo y de interpretación, no requiere muestras muy grandes y es más comprensible para personas con poco dominio de estadística (Holland & Thayer, 1988; Mellenbergh, 1982; Mazor, Clauser & Hambleton, 1992). Compara las respuestas (ciertas y erradas) de los grupos (en el caso, sexo masculino y femenino) en relación con un ítem. El nivel de habilidad de las personas que están siendo comparadas se obtiene por medio de la puntuación observada en el instrumento.

Finalmente, el modelo de Rasch (1960) parece ofrecer un repensar sobre el problema de la medida, al abordar la mayoría de los problemas en relación con las deficiencias del análisis tradicional de ítem, porque comenzó con suposiciones similares de medida (Rasch,

1960, 1966; Wright, 1968; Wright & Panchapakesan, 1969). Asumiendo la definición tradicional de sesgo del ítem, en el sentido de que un ítem es sesgado si mide de modo diferente distintos grupos, ese modelo evalúa tal efecto por medio del análisis de los residuos. Al lado de eso, como sus procedimientos son extensiones racionales del modelo, la investigación del sesgo puede ser más sistemática y más integrada (Birnbaum, 1968; Sisto, 2006).

RESULTADOS Y DISCUSIÓN

Inicialmente serán presentados los análisis referentes al TCT. El test *t* de Student fue utilizado para verificar la diferencia entre los sexos en los ítems del ECAP en los diferentes deciles. La tabla 1 ofrece los valores de *t* para los ítems que presentaron por lo menos un decil, una diferencia considerada significativa estadísticamente.

Como resultado de ese análisis, 19 ítems presentaron indicios de sesgo por lo menos en dos deciles. Específicamente, 12 ítems ofrecieron diferencias significativas en dos deciles, cuatro ítems en dos de ellos, un ítem en tres, otro ítem en cuatro y, finalmente, un ítem en siete de los deciles.

Tabla 1
Valores de t para los ítems y deciles respectivos

Ítem	Decil									
	1	2	3	4	5	6	7	8	9	10
1	-1.59	-1.34	-0.48	-0.90	-0.84	0.69	-1.66	-0.35	-2.07*	0.47
2	0.62	-0.12	0.36	0.89	-0.38	-1.24	0.18	0.63	0.07	-2.57*
3	0.46	-0.31	0.13	0.32	-0.99	-2.32*	0.30	-0.16	0.70	-1.33
8	-1.07	0.09	0.25	-1.04	0.42	-0.40	-0.18	2.54*	0.08	0.61
9	0.93	1.85	0.76	-0.50	-0.30	0.60	0.40	0.74	2.16*	-0.19
10	-1.92	1.12	1.30	0.50	2.53*	2.01*	-0.75	1.61	0.88	1.55
13	0.78	0.24	0.58	0.45	-1.34	1.14	1.44	-1.86	-2.05**	1.11
15	1.35	1.20	1.12	0.84	1.07	1.44	0.48	0.91	0.75	2.58*
16	-4.45***	-1.99*	-1.63	-2.39*	-2.26*	-2.92**	-2.72**	-3.23**	-1.03	-0.72
17	0.15	-0.85	-0.75	0.13	1.11	-3.04***	-0.83	-0.12	-0.79	1.73
18	-1.39	-1.80	-1.09	-0.32	-2.05*	-2.06*	-3.16**	-2.81***	-1.84	0.45
19	0.96	1.77	0.90	0.82	1.54	1.02	2.30*	1.55	0.20	0.96
20	0.99	1.82	0.06	0.30	1.58	1.29	2.15*	0.51	-0.03	0.71
29	0.63	1.53	1.40	1.08	3.08**	1.28	0.93	2.28*	1.96	-0.74
30	-0.83	0.14	3.28**	0.00	-0.93	-1.81	-1.19	-0.99	0.57	1.70
31	-1.17	-2.04*	-0.37	0.33	-1.76	0.22	-0.38	0.33	-2.13*	-1.16
32	-1.21	-2.08*	1.08	-1.21	-1.08	-0.22	-2.51*	-0.43	-0.17	0.93
34	-0.59	1.65	2.23*	2.03*	1.61	1.02	0.42	0.67	-1.11	2.70**
35	-1.10	1.72	1.15	2.94**	0.14	1.79	0.47	0.61	1.06	0.28

* p<0.05 ** p<0.001

A su vez, el procedimiento de Mantel Hantzel, por medio de la TRI, proporcionó datos de la tabla 2. Los resultados mostraron 17 ítems con funcionamiento diferenciado en razón del sexo de las personas.

Por fin, el análisis padrón realizado por Winsteps proporcionó los resultados constantes en la tabla 3. Por ese análisis 12 ítems acusaron favorecimiento en los aciertos y errores que privilegiaban un sexo u otro.

Tabla 2
Medidas de DIF, errores estándar, diferencias por sexos y tamaño de los valores de probabilidad de Mantel Hantzel para los ítems con diferencia significativa

Ítem	Sexo	Medida de DIF	Error estándar	Sexo	Medida de DIF	Diferencias	Mantel Hantzel prob.	Mantel Hantzel tamaño
1	M	52,29	0.35	F	51,05	-1.24	0,0142	-0.4
9	M	58,87	0.39	F	60,07	1.2	0,0396	1.94
10	M	46,44	0.45	F	48,11	1.67	0,0069	2.65
15	M	47,32	0.43	F	49,32	2.01	0,0014	0.61
16	M	50,06	0.38	F	45,13	-4.93	0	-6.51
18	M	48,66	0.4	F	45,95	-2.72	0	-3.08
19	M	46,66	0.45	F	49,06	2.4	0	2.22
20	M	46,39	0.45	F	48,25	1.86	0,0009	0.79
21	M	44,36	0.52	F	46,66	2.3	0,002	-2.16
22	M	55,26	0.35	F	54,48	-0.78	0,0208	0.46
26	M	55,21	0.35	F	53,82	-1.39	0,0279	-0.66
29	M	57,62	0.37	F	60,36	2.4	0	4.04
30	M	55,23	0.35	F	54,32	-0.91	0,0337	-1.19
31	M	50,02	0.38	F	49,23	-0.8	0,0348	-1.71
32	M	45,82	0.47	F	43,94	-1.88	0,0054	-3.63
34	M	49,93	0.38	F	51,31	1.38	0,0024	1.31
35	M	52,75	0.35	F	53,79	1.04	0,006	0.67

Tabla 3
Aumento del DIF, errores estándar, diferencias por sexo y valores de t para los ítems con diferencia significativa por el procedimiento Winsteps.

Item	Sexo	DIF aumentado	Error estándar	Sexo	DIF aumentado	Error estándar	Diferencia	Error estándar	t
1	F	-0.50	0.30	M	0.74	0.35	-1.24	0.46	-2,68
10	F	0.64	0.34	M	-1.03	0.45	1.67	0.56	2.97
15	F	0.76	0.32	M	-1.24	0.43	2.01	0.53	3.75
16	F	-2.27	0.40	M	2.65	0.38	-4.93	0.55	-8,95
18	F	-1.17	0.38	M	1.55	0.40	-2.72	0.55	-4,92
19	F	0.89	0.32	M	-1.51	0.45	2.40	0.55	4.37
20	F	0.70	0.33	M	-1.16	0.45	1.86	0.56	3.30
21	F	0.85	0.36	M	-1.46	0.52	2.30	0.63	3.63
26	F	-0.56	0.29	M	0.83	0.35	-1.39	0.46	-3,04
29	F	1.21	0.35	M	-1.53	0.37	2.74	0.51	5.35
32	F	-0.79	0.43	M	1.09	0.47	-1.88	0.64	-2,94
34	F	0.54	0.30	M	-0.84	0.38	1.38	0.48	2.87

Comparando los resultados de los tres procedimientos, el procedimiento padrón de Winsteps fue el que acusó un número menor de ítems con posibles funcionamientos diferentes en razón del sexo, en cuanto fue el modelo TCT el que acusó mayor número de ítems. Para averiguar las discrepancias entre los ítems clasificados con DIF por los tres modelos de análisis se construyeron

tablas de contingencia de dos entradas, en las cuales los ítems del instrumento fueron clasificados con y sin DIF. De esas tablas fueron calculados los chi cuadrados y derivadas las correlaciones por medio de los cálculos de los coeficientes de contingencia cuadrática media de Pearson y de correlación de Pearson. Los resultados se encuentran en la tabla 4.

Tabla 4
Ítems clasificados por presencia y ausencia del DIF por los 3 modelos, valores de χ^2 , grados de libertad, coeficiente de contingencia cuadrática media de Pearson y coeficiente de correlación de Pearson

		Winsteps			χ^2	gl	p	Coeficiente de contingencia	Coeficiente de Pearson
		Sem DIF	Com DIF	Total					
TCT	Sem DIF	14	2	16	6.21	1	0.013	0.39	0.42
	Com DIF	9	10	19					
Total		23	12	35					
		Winsteps			χ^2	gl	p	Coeficiente de contingencia	Coeficiente de Pearson
		Sem DIF	Com DIF	Total					
Mantel Hantzel	Sem DIF	18	0	18	19.34	1	0.000	0.60	0.74
	Com DIF	5	12	17					
Total		23	12	35					
		TCT			χ^2	gl	p	Coeficiente de contingencia	Coeficiente de Pearson
		Sem DIF	Com DIF	Total					
Mantel Hantzel	Sem DIF	13	5	18	10.49	1	0.001	0.48	0.55
	Com DIF	3	14	17					
Total		16	19	35					

Todos los resultados de los chi cuadrados informaron que las diferencias no podrían ser atribuidas al azar, a pesar de que en todas las tres comparaciones ha habido discrepancias. Los datos mostraron que de las 35 clasificaciones, comparando el modelo TCT con el padrón de Winsteps ocurrieron 11 discrepancias, Mantel-Hantzel con el padrón de Winsteps proporcionaron cinco discrepancias y, finalmente, con-

traponeo los datos oriundos del TCT y de Mantel-Hantzel fueron observadas ocho de ellas.

En razón de que esas diferencias no eran casuales, fueron derivadas de los índices de correlación, esto es, del coeficiente de correlación de Pearson y del coeficiente de contingencia cuadrática media de Pearson. Esos resultados también constan en la tabla 4 y sugieren una simetría de resultados en ambos proce-

Tabla 5
Ítems comunes a los tres modelos de análisis

Ítem	Descripción
Mayor frecuencia en el sexo masculino	
9	... ser subordinado
10	... comandar un equipo
15	... equipamientos
19	... elaborar proyectos
20	... ejecutar proyectos
29	... utilizar fuerza física
Mayor frecuencia en el sexo femenino	
1	... trabajar en lugares cerrados
16	... contacto con público/personas
18	... usar una expresión verbal

dimientos. Los criterios que mostraron una correlación más alta fueron los que usaron el modelo de Rasch, considerando que el modelo padrón de Winsteps apuntó el menor número de ítems con el DIF. La segunda correlación más alta fue entre los procedimientos del TCT y de Mantel-Hantzel.

A pesar de que esas correlaciones indicaron un grado de similitud entre los procedimientos, ellas también revelan que una buena parte de la varianza no se debe a esa comunalidad, mas sí a otros factores. En ese sentido, cabe realzar que los dos procedimientos con la mayor correlación apuntan a una comunalidad por vuelta de 55 %.

De los ítems con DIF en los tres procedimientos, los nueve ítems comu-

nes fueron separados por sexo. La mayor parte de ellos (seis ítems) facilitan conductas femeninas y los otros las masculinas.

Así, los resultados evidenciaron que las mujeres tuvieron medias mayores en ‘trabajar en lugares cerrados’, ‘atender públicos’ y ‘usar la expresión verbal’. Por su parte, los hombres escogieron más ‘ser subordinado’, ‘comandar un equipo’, ‘usar equipamientos’, ‘elaborar y ejecutar proyectos’ y ‘utilizar la fuerza física’.

CONSIDERACIONES FINALES

Este estudio pretendió analizar la diferencia entre hombres y mujeres en cuanto a las preferencias por ambien-

tes y características de trabajo con base en tres métodos diferenciados. Es sabido que hay diferencia entre preferencias profesionales (Cejka & Eagly, 1999; Lemos, 2001; Pérez & Jaramillo, 2007; Sartori, 2007; Sartori, Noronha & Nunes, 2009; Nunes & Noronha, 2009, entre otros), mas lo que justifica la realización de esta investigación es que ella aborda un instrumento en elaboración, cuya necesidad de análisis de variables distintas es permanente.

Así, a pesar de la diferencia entre los comportamientos de elegir entre sexos que ya ha sido objeto de investigaciones, parece pertinente que cada nuevo instrumento evalúe en qué medida se coloca esa diferencia. Eso se torna necesario para que se perfeccione la interpretabilidad de los instrumentos, no atribuyendo, por ejemplo, que las diferencias encontradas entre los sexos sea debida a esa variable, cuando la verdad es que los ítems del instrumento están facilitando la manifestación del constructo en personas de uno de los sexos en detrimento de otro. Bajo esa perspectiva, se debe acreditar que el objetivo fue atendido, que refuerza el imperativo del área de evaluación psicológica de que se elaboren nuevos instrumentos y que en relación, tanto con los nuevos como con los antiguos, se establezcan las cualidades psicométricas.

De manera especial, las discrepancias de los resultados de los distintos métodos permiten reflexiones. Por los modelos utilizados se llegaron a tres

conjuntos de datos distintos, aunque relacionados. Más específicamente, 19 ítems se comportaron de forma diferente para hombres y mujeres de acuerdo con la Teoría Clásica de los Test (TCT), 12 bajo la perspectiva de Winsteps y 17 del método de Mantel-Hantzel. De los análisis realizados, se llegó a un grupo de 9 ítems comunes, en los cuales la diferencia entre los sexos fue identificada. Así, fue posible afirmar que, indudablemente, por lo menos 9 actividades están facilitando las conductas masculinas y las conductas femeninas. A lado de eso, hay 10 actividades que podrían estar funcionando de manera diferente, favoreciendo a uno de los sexos. De hecho, la no concordancia es por lo menos igual a la concordancia, lo que quiere decir que hay un conjunto grande de ítems que están siendo indicados, por un método o por otro, como facilitando la conducta de determinado sexo.

Desde el punto de vista técnico está claro que al adoptarse un método u otro habrá siempre la posibilidad de error, pues ninguna correlación fue grande o suficiente como para decir que se trata de métodos que producen resultados paralelos. Entretanto, debe considerarse que elegir un método, a pesar de no proporcionar un criterio exento de falsos positivos y negativos, aun es mejor que no analizar o excluir ítems con posibilidades de funcionamiento diferente debido a variables que no miden lo que se quiere. De hecho, esta colocación implica que si aceptamos que una escala

exenta de DIF en la actualidad es bastante difícil de construir; sin embargo es posible minimizar bastante ese tipo de interferencia.

No obstante, se debe repensar sobre los modelos y las teorías usadas en los análisis aquí aplicados. A ese respecto, fue posible desprender que la Teoría Clásica de los Test está siendo (o ya fue) superada por otros modelos más modernos de análisis que, a su vez, utilizan procedimientos más esmerados. En ese contexto, Birnbaum (1968) destacó, hace más de cuatro décadas, que la dificultad del ítem en la TCT es dependiente de la muestra de acuerdo con la definición tradicional y que debe ser comprendida junto a un grupo en particular, pues cambia si fuera encontrado en grupos distintos, diferentemente del TRI, que postula una independencia de esa vinculación en sus modelos, en los que pese a tal consideración, no se puede, con todo, descartar las críticas referentes, por ejemplo, al modelo Mantel-Hantzel, dentro de ellas, de que posee poca potencia para detectar el DIF no uniforme (Mellenbergh, 1982).

Se debe también resaltar el hecho de que escoger el ECAP se debió a que se sabía que algunos de sus ítems necesitan tener un funcionamiento diferente, puesto que las profesiones, en cierta forma, poseen atractivos diferentes en relación con el sexo. Así, de los hallados pueden ser reforzados algunos ya identificados. Entretanto, en relación con el ECAP, específicamente Santos,

Noronha e Sisto (2008b), encontraron asociaciones entre los ítems ‘ser subordinado’, ‘comandar un equipo’, ‘usar equipamientos’ y ‘utilizar la fuerza física y las Ciencias Exactas’, identificada como una dimensión de la Escala de Aconsejamiento Profesional, cuyos hombres presentan medias mayores (Sartori, 2007; Sartori, Noronha & Nunes, 2009).

Al lado de eso, Santos, Noronha y Sisto (2008b) encontraron que ‘trabajar en lugares cerrados’, ‘usar la expresión verbal’ y ‘tener contacto con personas’ son actividades preferidas por las mujeres. De esa forma, se puede afirmar que hubo coherencia entre los resultados frecuentemente encontrados y los demás referenciados.

Se sugiere que se desarrollen nuevos estudios con los métodos en cuestión, más que evaluar constructos diferentes; asimismo se desea que haya en el ámbito brasileño mayor divulgación y utilización de las teorías modernas de análisis de datos.

REFERENCIAS

- Anastasi, A., & Urbina, S. (2000). *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F. Lord & M. Novick (Eds.), *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.

- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousands Oaks, CA: Sage Publications.
- Camilli, G. (1993). The case against item bias detection techniques based on internal criteria: Do item bias procedures obscure test fairness issues? En P. N. Holland & H. Wainer (Orgs.), *Differential item functioning* (pp. 397-413). Hillsdale, NJ: LEA.
- Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, 25, 413-423.
- Collier, C. M., Spokane, A. R., & Bazler, J. A. (1998). Appraising science career interests in adolescent girls and boys. *Journal of Career Assessment*, 6(1), 37-48.
- Conselho Federal de Psicologia (2009). Sistema de Avaliação dos Testes Psicológicos. Recuperado de <http://www.pol.org.br>
- Draba, R. E. (1977). The identification and interpretation of Item Bias. *Rasch Measurement Transactions*, MESA Memorandum 25. Recuperado de <http://www.rasch.org/rmt/rmt122m.htm>
- Holland, J. L., Johnston, J. A., & Asama, N. F. (1994). More evidence for the relationship between Holland's Personality Types and Personality Variables. *Journal of Career Assessment*, 2(4), 331-340.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. En H. Wainer & H. I. Braum (Orgs.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.
- Lemos, C. G. (2001). *Adolescência e escolha profissional*. São Paulo: Vetor.
- Lemos, C. G., Bueno, J. M. H., Balão, S. M. S., Silva, L. B., & Silva, P. L. (2005). Carreira profissional e relações de gênero: um estudo comparativo em estudantes universitários. *Boletim de Psicologia*, 55(123), 129-148.
- Lent, R., Lopez, S. D., & Bieschke, G. (1991). Toward a unifying social cognitive theory of career and academic interest, choice and performance. *Journal of Vocational Behavior*, 45, 79-122.
- Mazor, K. M., Clauser, B. E. & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement*, 52, 443-451.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, 7, 105-118.
- Melo-Silva, L. L., Noce, M. A., & Andrade, P. P. (2003). Interesses em

- adolescentes que procuram orientação profissional. *Psic*, 4(2), 6-17.
- Nunes, M. F. O., & Noronha, A. P. P. (2009). Interesses e personalidade: um estudo com adolescentes em orientação profissional. *Revista Galego-Portuguesa de Psicoloxía e Educación*, 17(1/2), 115-129.
- Pérez, E. R. & Jaramillo, V. A. (2007). Representaciones y acciones de los alumnos universitarios frente a la empleabilidad. *Psicología desde el Caribe*, 20, 130-155.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielson & Lydiche.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57.
- Rottinghaus, P. J., Lindley, L. D., Green, M. A., & Borgen, F. H. (2002). Educational aspirations: The contribution of personality, self-efficacy, and interests. *Journal of Vocational Behavior*, 61, 1-19.
- Santos, A. A. A., Noronha, A. P. P. & Sisto, F. F. (2008a). Interesses profissionais de adolescentes. Relatório técnico. Universidade São Francisco.
- Santos, A. A. A., Noronha, A. P. P., & Sisto, F. F. (2008b). Escala de características e ambientes profissionais. Relatório técnico não publicado.
- Sartori, F. A. (2007). *Evidências de validade de um instrumento de interesse profissional*. Disertación en la Maestría en Psicología. São Paulo: Universidade São Francisco.
- Sartori, F. A., Noronha, A. P. P., & Nunes, M. F. O. (2009). Comparações entre EAP e SDS: interesses profissionais em alunos do Ensino Médio. *Boletim de Psicologia*, LIX(130), 17-29.
- Sisto, F. F. (2006). O funcionamento diferencial dos itens. *Psico-USF*, 11(1), 35-43.
- Teixeira, M. A., Castro, G. D., & Cavalheiro, C. V. (2008). Escalas de Interesses Vocacionais (EIV): construção, validade fatorial e consistência interna. *Psicologia em Estudo*, 13(1), 179-186.
- Welter, G. M. R. (2007). O BBT-Teste de fotos de profissões em adultos e adolescentes. *Revista Brasileira de Orientação Profissional*, 8(1), 45-58.
- Wright, B. D. (1968). Sample-free test calibration and person measurement. En *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service.
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, 23-37.