

LA EVALUACIÓN DE LOS TRASTORNOS DE PERSONALIDAD EN LA INFANCIA

Evaluating personality disorders in childhood

Edgar Rodríguez Vílchez*

Resumen

El propósito del presente artículo consistió en investigar cualidades psicométricas para una prueba estandarizada que evalúe trastornos de personalidad en la infancia y adolescencia, para que de esta manera, pueda ser considerada como herramienta sólida que comprenda una confiabilidad y validez. Sin embargo, existen otras fuentes de evaluación, consideradas por los especialistas, que cumplen con los criterios para un diagnóstico de trastorno de personalidad. Por ello, para el diagnóstico y tratamiento del mencionado trastorno se recomienda tomar en cuenta un enfoque multifuente y multimétodo que incluya: entrevistas con el niño o adolescente y padres y/o cuidadores, administración de autoinforme, colección de la información colateral y revisión de los registros.

Se concluye que el diagnóstico de trastorno de personalidad en niños y adolescentes es novedoso y viable. Además, cabe mencionar que existe un porcentaje alto de investigaciones en problemas de personalidad en adolescentes, por ello es recomendable realizar más estudios en niños.

Palabras clave: Trastornos de personalidad, infancia, adolescencia, pruebas para evaluación, estandarización, diagnóstico.

Abstract

The purpose of this article was to investigate psychometric qualities for a standardized test to assess personality disorders in childhood and adolescence, so that in this way, can be considered as comprising a solid tool reliability and validity. However, there are other sources of evaluation, considered by specialists who meet the criteria for a diagnosis of personality disorder. Therefore, for the diagnosis and treatment of said disorder is recommended to take into account of a multisource and multimethod approach including: interviews with the child or young person and parents / carers, self management, collateral information collection and review of records.

It is concluded that the diagnosis of personality disorder in children and adolescents is novel and feasible. Furthermore, it is noteworthy that a high percentage of searches on adolescent personality problems, so it is recommended further studies in children.

Keywords: Personality disorders, childhood, adolescence, assessment testing, standardization, diagnosis.

* Doctor en Psicología en la Universidad Nacional Federico Villarreal en el año 2004. Master en Psicobiología de la Drogadicción en el año 1996. Docente de la Universidad Femenina del Sagrado Corazón. erodriguez_vilchez@hotmail.com

La evaluación juega un rol importante en todos los aspectos del trabajo clínico en niños y adolescentes que exhiben patrones de pensamiento y conducta trastornados. La evaluación proporciona la base para el diagnóstico inicial del trastorno de personalidad. Puede guiar también el desarrollo de un plan de tratamiento y puede ser utilizado para monitorear esfuerzos de intervención en el tiempo. Los procedimientos de evaluación pueden ser utilizados inicialmente para establecer patrones de pensamiento y comportamientos presentados por un niño o adolescente y establecer el grado según el cual estos patrones son juzgados para ser adaptables o maladaptativos. En el caso del trastorno de la personalidad, los procedimientos de evaluación se usan para proveer juicios sobre los comportamientos observados y, los patrones de pensamientos reflejan los importantes y perdurables problemas de ajuste. Cuando consideramos las necesidades del tratamiento de un niño o adolescente identificados como el tener un trastorno de personalidad, los resultados de la evaluación proporcionan la base para asignar el estatus de la clasificación administrativa para permitir el acceso a los servicios o para derivar al niño o adolescente a un proveedor de tratamiento apropiado. Una vez que se ha identificado el trastorno de personalidad, la evaluación puede proveer un marco para desarrollar un plan de tratamiento y para ayudar a los padres y a los profesores a entender y tratar con los problemas de comportamiento y los procesos de pensamiento trastornado del niño y el adolescente. Cuando el tratamiento progresa, los procedimientos de evaluación pueden permitir a los especialistas calibrar la eficacia de los esfuerzos del tratamiento.

Aunque el diagnóstico y el tratamiento de los trastornos de personalidad se ha restringido tradicionalmente a la edad de 18 años y más, el mayor sistema de clasificación importante usado por los especialistas, el manual estadístico y diagnóstico de los trastornos mentales (DSM-IV; Asociación Psiquiátrica Americana, 2000, p. 687), ha ofrecido algunas directrices para hacer un diagnóstico del trastorno de la personalidad en la niñez y la adolescencia.

Las categorías de los trastornos de la personalidad se pueden aplicar a niños o adolescentes en aquellas instancias relativamente inusuales en las que los rasgos de la personalidad maladaptativa particular del individuo aparecen siendo penetrantes, persistentes, y

poco probable que se limiten a una etapa de desarrollo particular o a un episodio de trastornos Eje I. Se debe reconocer que los rasgos del trastorno de personalidad que aparecen en la niñez frecuentemente no persisten sin modificaciones en la vida adulta. Para diagnosticar un trastorno de personalidad en una persona de 18 años de edad, las características deben haber estado presentes por lo menos 1 año.

Aunque en esta declaración el DSM-IV indica que el diagnóstico del trastorno de la personalidad en los niños y adolescentes debe ser una “instancia relativamente inusual”, la investigación resumió, en los capítulos que sigue, la prueba que sugiere que tales diagnósticos se justifican en un número mayor de instancias que actualmente declaradas. Como los mismos diagnósticos, la evaluación de los trastornos de la personalidad en los niños y adolescentes se encuentra en un campo emergente que no ha sido discutida de manera amplia o investigada en la literatura profesional. De hecho, no existen procedimientos generalmente aceptados para evaluar o diagnosticar los fenómenos de Eje II en niños y adolescentes (Westen, Dutra, & Shedler, 2005). Por ello, existe una necesidad de desarrollar métodos válidos y confiables que aseguren el tratamiento ético y humano de los niños y adolescentes.

¿EN QUÉ CONSISTE LA EVALUACIÓN DE LA PERSONALIDAD?

El significado del término “evaluación” varía ampliamente en la literatura de la psicología profesional, especialmente en tanto es aplicado a las áreas de los trastornos de la personalidad. Las definiciones de la evaluación de la personalidad van desde las concepciones muy amplias como las de Wiggins (2003) hasta las delimitaciones muy específicas como las de Hilsenroth y Segal (2004). En este extenso resumen del campo titulado Paradigmas de la Evaluación de la Personalidad, Wiggins define la evaluación de la personalidad como un proceso que involucra recopilar y evaluar varias fuentes de información acerca de un individuo para servir a propósitos tales como la selección, disposición, diagnóstico y formulación de caso. Dentro de este amplio marco, los métodos específicos de evaluación pueden incluir las pruebas psicológicas, entrevistas, y material biográfico. Por el contrario, Hilsenroth y Segal tienen una perspectiva mucho más limitada respecto

a qué constituye la evaluación de la personalidad, como se evidencia en su selección de contenido para la Evaluación de la Personalidad, volumen 2 de su *Comprehensive Handbook of Psychological Assessment* (Manual Completo de Evaluación de Psicológica). En esta obra de referencia, la evaluación de la personalidad se refiere a las pruebas específicas de la personalidad dividida en dos amplias categorías: Las evaluaciones objetivas y las evaluaciones proyectivas. Dentro de este marco, las evaluaciones objetivas incluyen pruebas objetivas diseñadas para evaluar los constructos de la personalidad o los trastornos psiquiátricos así como técnicas de entrevistas estructuradas y semiestructuradas debido a su enfoque estandarizado para la recopilación y evaluación de información. La definición de las evaluaciones proyectivas dentro de este marco es algo menos específico, sin embargo la dimensión unificadora es el hecho de que en el rango de las técnicas proyectivas discutidas, cada una proporciona una estructura de condiciones similares para la recopilación, evaluación, e interpretación de las muestras de comportamientos.

Como lo señaló Wiggings (2003), las conceptualizaciones de la personalidad carecen de valor científico a menos que se puedan operacionalizar y convertir en procedimientos de medición concretas. A los efectos de la práctica clínica, debe haber un marco formal y una serie de procedimientos para la emisión de un juicio acerca de la adecuación de estos procedimientos de medición operacionalizada. Se establece un marco aprobado de manera profesional y una serie de procedimientos para este propósito en *The Standards for Educational and Psychological Testing* (Normas para las pruebas psicológicas y de educación), conjuntamente desarrollado por *The American Educational Research Association* (Asociación Americana de Investigación Educativa, AREA por sus siglas en inglés), *The American Psychological Association* (Asociación Americana de Psicología, APA por sus siglas en inglés), y *the National Council of Measurement in Education* (Consejo Nacional de Medición en educación, NCME por sus siglas en inglés). Los estándares direccionan específicamente temas como los procedimientos de estandarización y el establecimiento de la confiabilidad y la validez, y establecen directrices para los métodos y procedimientos aceptables para demostrar la adecuación psicométrica de un procedimiento de evaluación. Ya que las consideraciones de la adecuación

de amplio rango de enfoques de evaluación discutidos por Wiggings (2003) está yendo más allá del alcance de este capítulo, se hace hincapié aquí en la discusión de las consideraciones psicométricas para el desarrollo y uso de las evaluaciones de los trastornos de la personalidad en el contexto de la definición utilizado por Hilsenroth y Segal (2004), especificando las evaluaciones de la personalidad como pruebas objetivas (que incluyen entrevistas estructuradas y semiestructuradas) y técnicas proyectivas. Ya que este capítulo aborda la evaluación de las personalidades de los niños y adolescentes, se incluye otra importante forma de evaluación en la categoría de prueba objetiva: Escalas de clasificación basadas en los informes de los otros como los padres, cuidadores y profesores. Para fines de este capítulo, el término “prueba” frecuentemente se usa cuando se hace declaraciones globales que se supone se refieren a todas las formas de evaluación de personalidad abarcadas por la definición de Hilsenroth y Segal.

Cuando las escalas de clasificación y de autoinforme se designan como un solo grupo, se denominan como “medidas de autoinforme y de informe a otros” o simplemente como pruebas objetivas. Los lectores se interesan en considerar la naturaleza de las técnicas de la evaluación desde una perspectiva más amplia que la presentada aquí referida a los *Paradigms of Personality Assessment* (Paradigmas de Evaluación de la Personalidad) de Wiggings como un punto de partida potencial.

CONSIDERACIONES PSICOMÉTRICAS EN EL DESARROLLO Y USO DE LAS PRUEBAS DE LA PERSONALIDAD

Los métodos de evaluación de la personalidad varían ampliamente en las orientaciones teóricas de los proponentes de los métodos y en la manera en la cual las concepciones teóricas se operacionalizan para formar instrumentos de pruebas. Independientemente de la orientación teórica de los autores de las pruebas o el formato específico de una prueba, los especialistas deben demostrar ciertas cualidades psicométricas básicas para el uso de las pruebas profesionalmente sólidas. Las siguientes secciones discuten estas propiedades psicométricas básicas bajo el título de “estandarización, confiabilidad y validez”. La estandarización se refiere a los procesos envueltos en la construcción y desarrollo de las pruebas que incluyen

ítems de desarrollo, administración, procedimientos de conteo, y marcos de interpretación de puntuación.

La confiabilidad se refiere a la consistencia y estabilidad de los resultados obtenidos en las pruebas. La validez se refiere a la evidencia proporcionada para demostrar que las pruebas miden lo que se planea medir y que la prueba se puede usar para sus propósitos establecidos. Aunque estas características son discutidas en forma separada, son elementos entretejidos inextricablemente. Cuando se hace de forma correcta, el proceso de construcción de una prueba por este mismo es una forma de validación de la prueba, y aunque para una prueba es teóricamente posible ser validado sin ser confiable, los instrumentos de evaluación sólidos deben demostrar un grado adecuado de confiabilidad antes de que se puedan utilizar de manera válida para asistir a la formulación de los juicios profesionales.

ESTANDARIZACIÓN

Durante la fase de desarrollo, la selección del contenido de las pruebas se determina por la orientación teórica específica de los autores de las pruebas.

A pesar de las declaraciones por parte de algunos autores de las pruebas quienes aseguran una posición ateorética, todas las pruebas tienen una base teórica. Lo que varía no es el grado por el cual se basa una prueba en la teoría, sino el grado en el cual los autores de la prueba admiten conscientemente los fundamentos teóricos del contenido de las pruebas. Aunque la teoría y los propósitos previstos de la prueba impulsan la generación del contenido de la prueba, en el análisis final estos aspectos de construcción de la prueba llegan a ser temas de validez. Más allá de lo racional en lo que respecta a la selección del contenido, la estandarización de una prueba implica cómo las ideas se convierten en concreto, medidas operacionalizadas, es decir, cómo las ideas acerca del contenido de las pruebas se traducen en ítems de prueba, cómo se administra y califica los ítems de prueba, cómo se especifican los marcos de referencia (generalmente en la forma de normas de pruebas) para la interpretación de la puntuación. Se debe garantizar que la prueba se construye de manera que permita la puntuación y la administración consistente a través del uso de materiales y de formatos de elementos bien diseñados y fácilmente utilizados que son consistentes con los previstos propósitos de evaluación del instrumento.

CONFIABILIDAD

La confiabilidad definida en relación a las pruebas de personalidad, se refiere a la consistencia de los resultados de las pruebas de la personalidad. Mientras más confiable es una prueba, es más probable que la persona examinada reciba la misma, o casi la misma puntuación en las repetidas administraciones de la prueba. Una prueba de la personalidad intenta averiguar las características de la personalidad de un individuo a través de las respuestas de la persona a los ítems de prueba, o a través de las respuestas proporcionadas por otros que conocen a la persona. En el caso de las pruebas objetivas, las respuestas de los ítems reflejan los juicios basados en las recolecciones y reflexión de las acciones, sentimientos, pensamientos y percepciones seleccionados que han ocurrido, o se creía ocurrido, durante un periodo de tiempo determinado y no determinado. La confiabilidad de las pruebas objetivas, por lo tanto, indica la consistencia de las percepciones de los encuestadores de la prueba. En el caso de las pruebas proyectivas, se considera que las respuestas de los ítems son representaciones significativas que son interpretadas por el examinador reflejando maneras que la persona tiende a percibir, pensar, sentir y actuar. Por ello, la confiabilidad de las pruebas proyectivas indica la consistencia de las representaciones significativas proporcionadas por la persona que responde, así como la consistencia de las interpretaciones atribuidas a aquellas representaciones a través del juicio del examinador.

Entender que la confiabilidad es una función de los juicios de la persona que responde o de las producciones del que responde y los juicios del examinador, debería ser evidente que varios factores influyen enormemente en la confiabilidad de una prueba, que incluyen características de la prueba, características de los médicos quienes dan las pruebas y, las características de los encuestados que toman la prueba. Metodológicamente, existen dos enfoques básicos usados para establecer la fiabilidad de una prueba: métodos de test-retest (prueba y re prueba) y los métodos de consistencia interna. Un tercer método, la confiabilidad entre los evaluadores, se puede usar para describir la consistencia de los evaluadores en sus aplicaciones de los criterios de puntaje de evaluación.

Características de las pruebas

La consistencia de los resultados de las pruebas puede estar enormemente influenciada por varios

factores relacionados con la construcción de la prueba. Estos incluyen el desarrollo de las direcciones de administración de la prueba, los formatos de los ítems y, las reglas de puntuación de los ítems. Tanto para ambas pruebas, las objetivas como las proyectivas, el desarrollo de las direcciones de administración de la prueba se debe dar cuidadosamente para asegurar que todos los evaluadores y los evaluados, en la mayor medida posible, tengan el mismo conocimiento de la naturaleza de la tarea y cómo esto se va a lograr. Las declaraciones a los evaluadores deben explicar claramente lo que se ha dicho a los evaluadores y cómo están siendo explicadas las tareas o los ejemplos. Las declaraciones hechas a los examinados por los evaluadores deben ser redactadas con tanta claridad como sea posible de manera que no haya confusión acerca de lo que se espera en términos de desempeño. Las instrucciones pobremente especificadas del evaluador y las direcciones del evaluado incrementan la posibilidad de que los evaluadores alteren la manera en el cual ellos comunican la naturaleza de la tarea y, consecuentemente los examinados puedan cumplir con las tareas de manera diferente de un momento a otro.

La manera en que los ítems de las pruebas son formateados y presentados también puede afectar el desempeño de la persona evaluada. En el caso de las pruebas objetivas, la redacción de los ítems debe recibir una cuidadosa revisión del contenido profesional y editorial para asegurar que son fáciles de leer y entender. Las construcciones gramaticales complejas y las dobles negativas deben evitarse. Por ejemplo, la declaración “me molesto mucho que no puedo controlarlo” se prefiere en mayor medida que la declaración “no me molesto mucho que no puedo controlarlo”; Es mucho más probable que la primera declaración sea comprendida y completada en una manera coherente con las respuestas previstas del evaluado. Se debe tener cuidado a la hora de la selección de las descripciones de la opción de respuesta cuando usemos las escalas multipunto. Cuanto más claramente definida esté cada opción de respuesta, mayor probabilidad que un evaluado responda consistentemente al ítem en más de una ocasión. Cuando las opciones no se encuentran claramente definidas, el examinado o evaluado debe confiar en su propio juicio perceptivo para hacer distinciones entre las opciones. De ese modo se incrementa la probabilidad de que un evaluado interprete las opciones

en una manera diferente y de forma impredecible si se toma la prueba más de una vez.

La disposición y el diseño del formulario de prueba deben estar bien sopesados. Los formularios de prueba que firmemente apiñaron las preguntas en un espacio pequeño y usaron un tipo pequeño corren el riesgo de que los evaluados omitan los ítems o desalineen respuestas con los ítems. Los formularios que emplean múltiples columnas en una sola página debe también diseñarse cuidadosamente para evitar confusión en relación a donde deben registrarse las respuestas. Para las pruebas proyectivas, el formato de los ítems de pruebas normalmente no es un problema. Las pruebas de frases incompletas tienden a ser inequívocas y a ser bien estructuradas en términos de formato. Las tareas como las de Rorschach y la Prueba de Apercepción Temática usan una tarjeta para cada ítem. Las medidas gráficas no tienen ítems para formatear.

Características del Examinador

La manera en que el examinador se compromete en la administración de la prueba y la puntuación puede afectar en gran medida la consistencia de los resultados de las pruebas. Se debe tener mayor cuidado en desarrollar de manera correcta la estandarización de las instrucciones de administración del examinador o evaluador y las direcciones del evaluado. Igual cuidado debe tomarse por parte del examinador para asegurarse que las instrucciones y las direcciones sean entregadas de un modo que no conduzca a un desempeño inconsistente. El cumplimiento estricto de las instrucciones estandarizadas y las direcciones es crucial en todas excepto en la más inusual de las circunstancias. Las direcciones saltantes, reducciones, y reelaboración pueden reducir la probabilidad de que los ítems puedan ser completados de forma consistente. Las interacciones pueden también afectar el desempeño de la persona que responde. Un comportamiento profesional propio durante las presentaciones de las instrucciones y las direcciones comunica una actitud más sensata hacia la situación de prueba y maximiza la probabilidad que el evaluado se acerque a la prueba de manera igualmente seria. La mayor amenaza de la confiabilidad de la perspectiva del examinador o evaluador, y quizás de todas las perspectivas, es una falta de adherencia a la aplicación estandarizada de los ítems de los reglamentos de la puntuación. Esta amenaza es especialmente evidente en el caso de las pruebas proyectivas, donde los sistemas de puntuación

complicados deben ser aprendidos y aplicados con la confidencialidad para asegurar resultados confiables. Con la excepción de los formularios de escáneres electrónicos usados con algunas pruebas objetivas, la necesidad de realizar un simple cálculo y tareas de oficina para obtener la primera puntuación calculada y la puntuación de conversión basadas en normas está presente en algunos grados con cada administración de una prueba de personalidad. El cálculo básico y los errores de puntuación de la oficina son una mayor fuente de prueba de una prueba de no confiabilidad y, uno que a menudo no se refleja en los estimados cuantificados de las pruebas de confiabilidad reportados en los manuales de pruebas, como en los casos usados en los estudios de confiabilidad que normalmente y en repetidas ocasiones se chequean minuciosamente por posibles errores de puntuación.

Características de los evaluados

No es importante qué tan bien se construya una prueba o cuán competente sea el examinador o evaluador, quien lo administra, y completa sea esta, los estimados de confiabilidad de una prueba nunca reflejarán una consistencia perfecta. La naturaleza inalcanzable de la consistencia perfecta se debe en gran parte a la tercera fuente de influencia de confiabilidad de la prueba: los evaluados quienes toman la prueba. Aun cuando las personas que responden están determinadas a realizarlo de manera consistente, las dos características de la naturaleza humana hacen que se interpongan en el camino de una consistencia perfecta, y todos los humanos perciban, piensen, sientan y, actúen de forma diferente en cualquier momento dado, e incluso en la presencia del mismo estímulo. Las tendencias de los evaluados hacia la inconsistencia afecta el desempeño de la prueba objetiva como ellos afectan en el desempeño de la prueba proyectiva. En el caso de las pruebas objetivas, las variaciones biológicas en las capacidades humanas aseguran que algunas personas que responden a las pruebas no estarán atendiendo de la misma manera a todas las direcciones todo el tiempo que ellos tomen las pruebas, tampoco atenderán algunos o tratarán igual y efectivamente cada ítem en la prueba. Tales lapsos de atención y esfuerzos del proceso resultan en patrones de respuesta inconsistente a través de administraciones de pruebas múltiples. Algunos encuestados no percibirán en gran número y de la misma forma las demandas de las tareas cuando hablan de la prueba

más de una vez, de este modo alterarán su desempeño a través de la administración. Algunos de los que responden encontrarán difícil interpretar de manera consistente la diferencia entre las opciones multipunto de respuesta, no importa qué tan claro parece ser la diferenciación entre ellos, cambia de uno a otro a través de las administraciones (ejemplo, cambiando de “Casi todo el tiempo a Algunos”). Algunos encontrarán difícil interpretar consistentemente la diferencia entre elecciones dicotómicas tales como “verdad” “falso” cuando consideren las declaraciones acerca de cómo perciben, piensan, sienten y actúan.

La última dificultad podría parecer excepcionalmente caprichosa en tanto existan solo dos elecciones a considerar, sin embargo en la actualidad, la dificultad con las opciones de respuesta dicotómicas es quizás la más entendible. Si se enfrentan con la elección verdad-falsa, cómo debería un adolescente responder razonablemente a la declaración siguiente: “¿mis padres controlan mi vida?”, qué pasa si, como los adolescente reflexionan en los eventos de los pasados 6 meses, él o ella siente que algunas veces los padres ejercen una cantidad no razonable de la influencia, pero en otros tiempos ¿estuvieron los padres extremadamente abiertos para permitir al adolescente hacer lo que él o ella elija? Cómo entonces, ¿debería responderse la pregunta: verdadero o falso? Uno puede fácilmente imaginar un escenario en el cual, habiendo sido ya confrontado acerca de permanecer muy tarde la noche anterior, un adolescente evaluado elige la “verdad” aunque al mismo tiempo él ha recordado muchas veces cuando los padres no ejercían tal control. La próxima vez que el adolescente tome la prueba, ningún evento negativo como tal habrá ocurrido recientemente. Ahora, sin una restricción y sintiéndose muy bien con respecto a los padres, el adolescente da mayor importancia a todos los casos en la cual le permiten elegir y responder con un “falso”. Aunque los evaluados son conducidos a considerar sus percepciones, pensamientos, sentimientos y acciones a través de un periodo de tiempo extendido, es difícil aun para los evaluados muy bien entrenados, dejar solo a un niño o adolescente promedio, para participar en el recuerdo objetivo que igualmente da importancia a todas las circunstancias destacadas del pasado (y solo circunstancias sobresalientes) para llegar a una estimación cuantitativa de la frecuencia de ocurrencia para una declaración específica.

Una discusión previa se centra en los autoinformes acerca de sus percepciones, pensamientos, sentimientos y acciones, incluso en cuestiones planeadas que se aplican de igual forma a las clasificaciones proporcionadas por otras personas quienes están familiarizadas con las percepciones, pensamientos, sentimientos y acciones del paciente como los miembros de la familia (especialmente los padres) y profesores. Las clasificaciones de estos individuos a menudo se usan cuando tratan de determinar las características de la personalidad de los niños y adolescentes. También se usan algunas veces con adultos cuando el paciente no es visto como una fuente fiable de información de autoinformes. Las escalas de clasificación de este tipo pueden asegurar una descripción más precisa de la personalidad de un niño o adolescente, pero como las pruebas mismas, su confiabilidad debe también demostrarse.

El tema de la capacidad para el recuerdo balanceado de experiencias pasadas trae a la luz otro factor del evaluado que puede afectar la consistencia de la respuesta: la edad del encuestado. La inspección de la prueba y la reprobación y la información de confiabilidad de consistencia interna reportado en varios manuales de prueba de personalidad objetiva muestra un patrón consistente a través del cual la estimación de confiabilidad incrementa paulatinamente con la edad de los evaluados.

A diferencia con la progresión vista en los manuales de la prueba de personalidad, la prueba y la reprobación y la información de consistencia interna proporcionada en los manuales de prueba de evaluación e inteligencia no muestran un patrón similar de progresión, como las estimaciones de confiabilidad que permanecen muy parecidas en todas las edades entre 8 a 90 para las tareas de contenido similar. Aunque se podría plantear la hipótesis que la razón para las diferencias de confiabilidad es que las pruebas de personalidad adulta están mejor construidas, tal hipótesis parece poco probable a la luz del hecho que las revisiones profesionales proporcionan evidencia para apoyar la premisa que la mayoría de pruebas de la personalidad del niño y adolescente están tan bien construidas como las mediciones adultas. Una explicación más probable para el efecto de la progresión de la confiabilidad es que las medidas de autoinforme de la personalidad colocan las demandas del proceso cognitivo específico en los evaluados, y estas demandas de procesos son

mucho más sensitivos a la maduración de capacidades neurales que otros tipos de pruebas. Las evaluaciones de autoinforme de la personalidad normalmente no se estandarizan con las muestras de mucho más jóvenes de 8 años. Cumplir los 8 años de edad, sin embargo, no significa que los niños ahora tienen la misma capacidad como un adulto, o incluso un adolescente, para responder de manera consistente a las preguntas que las requieren para reflejar y emitir juicios acerca de sus propias percepciones, pensamientos, sentimientos y acciones en una manera objetiva. Tal capacidad probablemente lo desarrolle lentamente con el tiempo, acorde con la maduración de las áreas específicas del cortex prefrontal que desempeña un papel importante en desarrollo de la autoreflexión y capacidades de juicio.

La tendencia del desarrollo observado con las pruebas objetivas a menudo no se presenta en la información de confiabilidad reportada por las pruebas proyectivas de la personalidad. En el caso de las pruebas proyectivas, aunque la naturaleza del desarrollo de las producciones del evaluado es aparente (Leichtman, 2004), los estudios repetidos de 3 semanas de administraciones de pruebas proyectivas a los niños han reportado estimaciones de confiabilidad acorde con aquellas obtenidas por las muestras de los adultos (Weiner, 2004). Estos resultados son comprensibles ya que en los casos de las pruebas proyectivas las percepciones de los evaluados producen directamente la información que son la base de las estimaciones de confiabilidad. Por el contrario, en el caso de las pruebas objetivas, las producciones de los evaluados reciben una puntuación por los evaluadores, y estas puntuaciones, producidas por los evaluadores y no por los evaluados, son la información que es la base de las estimaciones de confiabilidad. Al reconocer la naturaleza del desarrollo de las producciones de pruebas proyectivas del evaluado, los que desarrollan la prueba proporcionan un sistema de puntuación que controla las variaciones de desarrollo identificable en respuesta. Lo ilustrativo de la naturaleza sobre el control de las influencias de desarrollo en los sistemas de puntuación de pruebas proyectivas, es el hecho que una imagen de una persona, o un respuesta Rorschach específica, no sería calificada de la misma manera o recibiría la misma puntuación cuantitativa cuando es producida por un niño de 8 años de edad o un adulto de 45 años de edad. Aunque, estos sistemas de puntuación de desarrollo controlado por la variación de

desarrollo en el tiempo, no necesariamente controlan cómo un evaluado elige expresar sus percepciones, pensamientos, sentimientos y tendencias para actuar a través de la producción de una respuesta al estímulo ambiguo de una prueba proyectiva. La variación en las producciones del evaluado de una administración a otra, consecuentemente, permanece una fuente potencial de inconsistencia en la medida de confiabilidad. Las pruebas objetivas se encuentran en un mayor riesgo de la influencia de la respuesta inconsistente debido al estatus de desarrollo del evaluado, y las pruebas proyectivas son más vulnerables a las amenazas a la confiabilidad de las fuentes combinadas de la inconsistente producción del evaluador y la incongruente puntuación del evaluador. El hecho de que las puntuaciones de las pruebas proyectivas se encuentran tan confiables como la mayoría de las otras formas de medida (Meyer, 2004) es un crédito para los desarrolladores de prueba quienes ponen un gran esfuerzo para estandarizar la administración de pruebas proyectivas y calificar los estándares más altos posibles.

CONSISTENCIA INTERNA DE LA ESCALA

La metodología de la consistencia interna proporciona estimaciones de la extensión en la cual las respuestas a los ítems de una prueba, o las subsecciones de la prueba, son consistentes con cada otro. Tales estimaciones se pueden calcular en los contenidos enteros de una prueba de personalidad, pero ellos normalmente se calculan en los subconjuntos de los ítems, usualmente referidos a tales escalas, las cuales a menudo se dividen a su vez en subgrupos de subgrupos, referidos a tales subescalas. La composición de la escala y la subescala de una prueba representa lo operacionalmente definido relaciones estructurales de los constructos de personalidad que la prueba pretende medir. Por ejemplo, una prueba objetiva de trastornos de internalización se podría componer de dos escalas etiquetadas como Depresión y Ansiedad. Estas dos escalas se pueden dividir a su vez en tres mediciones de subescalas que representan los componentes cognitivo, afectivo y psicológico, para un total de seis subescalas. Cada una de estas seis subescalas, así como las dos escalas posiblemente sería analizada por la consistencia interna. El tipo más común de análisis de consistencia interna produce una correlación coeficiente llamada "coeficiente alfa", el cual representa un promedio de la relación de cada ítem en la escala o subescala

con cada otro ítem en la escala o subescala. Estas estimaciones pueden extenderse de 0.00 a 1.00, con enormes valores que indican grandes consistencias entre los ítems. Es conveniente un grado razonable de consistencia, ya que indica que un evaluado está proporcionando respuestas que conceptualmente van juntas a representar un patrón de percepción, pensamiento, sentimiento o acción. Por ejemplo; si una subescala de la prueba objetiva compuesta de 10 ítems pensado en medir las características de la personalidad borderline produce una estimación de coeficiente alfa de 0.20, este valor sugiere que una mayoría de los evaluados no respondieron consistentemente a los 10 ítems. La falta de consistencia podría ser debido a la formulación o puntuación de algunos aspectos de los ítems o de las opciones de respuestas, o debido a una falta de coherencia conceptual entre las percepciones, pensamientos, sentimientos, o acciones representadas por los ítems. Cualquiera de estas fuentes sugeriría que algunos aspectos de la construcción de la prueba no fueron completados lo suficientemente, y adecuados para producir una medida consistente.

En el caso de las pruebas proyectivas, las puntuaciones del evaluador para un número de características de una producción del evaluado podrían ser agregadas para producir un puntaje de índice. Si la estimación de consistencia interna para el agregado de estas características de puntuación fuera de 0.20, este valor sugeriría que una mayoría de examinadores no asignaron el puntaje de la misma o similar magnitud de cada una de las características. La falta de consistencia podría ser debida a una falta de claridad en las direcciones de puntuación, de la incapacidad de los varios examinadores para juzgar objetivamente las características, de la inconsistencia en la manera que los evaluados producen las características, o de una falta de lo conceptual que encajan con estas características como un grupo para representar una constructo cuantificable. Cualquiera de estas fuentes sugeriría que algunos aspectos de la construcción de la prueba no fueron completados lo suficientemente, y adecuados para producir una medida consistente.

La premisa fundamental de la construcción de escalas que usa múltiples ítems para evaluar los constructos psicológicos es que cada ítem, mientras es representativo del completo constructo, también representa un único aspecto del constructo que no es necesariamente capturado por algún otro ítem

en la escala. Los constructos complejos requieren varios ítems para la evaluación porque pocas personas probablemente demuestren todo de las características que son representativas del constructo, pero probablemente manifiesten un gran número de todas las posibles características. Esta premisa está en cierto modo contrarrestando la idea de la consistencia interna en que si los ítems de una escala fueran consistente interna y completamente, significaría que los evaluados estuvieron respondiendo de la misma manera a cada ítem de la escala. Cada persona sería una pura representación del constructo o no. Si ese fuera el caso, no habría necesidad de incluir múltiples ítems en la escala, un solo ítem bastaría para una medida precisa. Ya que el escrito puro de este tipo no es una caracterización precisa de cómo las personas manifiestan los constructos psicológicos, las estimaciones de consistencia interna extremadamente altas no son realmente deseables en un prueba de personalidad. Valores moderados en lo alto .70s a lo alto.80s son ciertamente adecuados para indicar un grado lo suficientemente alto de la consistencia interna.

Los juicios finales de la confiabilidad de la prueba se debe hacer solo después de considerar las fuentes adicionales de información acerca de la consistencia de la medida, incluyendo la confiabilidad de la prueba- re prueba y, si es dictado por el tipo de prueba, confiabilidad entre evaluadores.

LA CONFIABILIDAD ENTRE EVALUADORES

Cuando los juicios del evaluador juegan un rol crucial en la puntuación de una prueba, como es el caso para las pruebas de personalidad proyectiva, todas de manera virtual, es necesario determinar el grado de consistencia del puntaje proporcionado por los evaluadores. Los métodos usados para cuantificar la consistencia del puntaje del evaluador se refieren a los análisis de la confiabilidad entre los evaluadores. Estos análisis normalmente se conducen usando métodos de correlación similar a aquellos resultados predichos. En lugar de correlacionar las puntuaciones de los evaluados a través de dos administraciones de la misma prueba, el análisis de la confiabilidad entre evaluadores correlaciona las puntuaciones asignadas por dos o más evaluadores a las respuestas de las pruebas de un grupo de evaluados. El acuerdo de los evaluadores se afecta en su mayoría por dos factores:

el sistema de puntuación aplicada para obtener puntuaciones (una característica de la prueba) y el juicio y la habilidad profesional de los evaluadores. Ya que el mismo conjunto de respuestas o producciones obtienen puntuaciones por los evaluados, las características de las respuestas no tienen ningún efecto en la consistencia de las puntuaciones derivadas de ellos.

Como ya se ha discutido en detalle, el desarrollo de los sistemas de puntuación bien concebido y bien estandarizado es fundamental para la evaluación confiable. Cuando un gran grupo de evaluadores capaces no pueden demostrar la puntuación consistente, es bastante probable que haya una falla en el sistema de puntuación que necesite ser direccionado para mejorar la confiabilidad de la prueba. En cambio, cuando un grupo de evaluadores novatos pueden producir grandes resultados consistentes, es bastante probable que el sistema de puntuación sea sólido y es probable que demuestre un alto grado de consistencia para los evaluadores con distintos niveles de experiencia.

En casos donde un grupo de novatos evaluadores falla en producir resultados de puntuación consistentes, es difícil saber si el deficiente acuerdo se dio debido a la inexperiencia del evaluador o a las guías de puntaje deficiente. Ya que estas dos fuentes de variabilidad están siempre potencialmente presentes en algún grado, es importante que el análisis de confiabilidad direcciona el efecto de la experiencia del examinador en la confiabilidad de la prueba. Esto se puede dar a través del estudio separado hechos con los novatos y con los evaluadores experimentados, o puede darse a través de la inclusión de la experiencia del examinador como una variable moderadora en el análisis de la data de confiabilidad.

LA VALIDEZ

Aunque la validez de la prueba ha sido un concepto fundamental en el desarrollo de la prueba desde su implantación, el concepto ha experimentado una transformación substancial en las últimas décadas (Goodwin, 2002; Goodwin & Leech, 2003).

El actual Standards for Educational and Psychological Testing (Estándar para la Prueba Educativa y Psicológica) emitió en 1999 la validez establecida y el proceso de validación de una prueba:

La validez se refiere al grado por el cual la evidencia y la teoría respaldan las interpretaciones de las pruebas conllevadas por los usos de las pruebas propuestas. Por ello, la validez es la consideración fundamental en el desarrollo y evaluación de las pruebas. El proceso de validación incluye la acumulación de evidencia para proveer una base científica sólida para las interpretaciones de puntuación propuesta. Esto es, las interpretaciones de las puntuaciones de las pruebas requeridas por el uso propuesto que son evaluados, no por la misma prueba. Cuando las puntuaciones de las pruebas se usan o se interpretan en más de una forma, cada interpretación debe ser validada (American Educational Research Association [AERA], American Psychological Association [APA], y National Council on Measurement in Education [NCME], 1999, p.9).

La validez de una prueba de personalidad relacionada con la adecuación de las evidencias reunidas para respaldar los usos profesionales previstos de las pruebas. Las pruebas que carecen de validez son consideradas fuentes deficientes de información acerca de los evaluados, ya que se piensan en ellos dando poco en la manera de una caracterización precisa de rasgos de personalidad y poco en la manera de información útil que se pueda usar para ayudar a la toma de decisiones profesionales acerca de la situación de un paciente. En cambio, una prueba válida es una que ha demostrado evidencia de la efectividad de la prueba caracterizando de manera precisa los rasgos de la personalidad y proporcionando información valiosa que mejora la toma de decisión clínica.

El proceso de validación de una prueba de personalidad incluye la reunión de la evidencia para apoyar la efectividad. Los autores de la prueba y los editores resumen la evidencia que han acumulado para respaldar una validación de una prueba en una o más capítulos en el manual de la prueba. Estos hallazgos normalmente se presentan usando un tradicional, pero no anticuado, formato que ha utilizado tres títulos distintos -validez del contenido, validez relacionada al criterio, y la validez de constructo- y tratan estas tres áreas como si fueran tipos distintos de validez. La validez del contenido normalmente se refiere a la adecuación del contenido de la prueba para representar los constructos que se piensan son medidos por la prueba. La validez relacionada con el criterio se refiere a la comparación de los resultados de las pruebas con otras medidas del mismo constructo o con los

resultados previstos. La validez de constructo se ha referido a lo extenso por el cual la prueba mide los rasgos o el constructo que pretenden medir.

Admitiendo la superposición sustancial de estos tres tipos de validez tradicional y sus carencias de ajuste a las concepciones actuales de validez, los Standards for Educational and Psychological Testing presentaron una reconceptualización de validez, distinguiendo entre los tipos de información que deben ser recogidos para demostrar la validez de una prueba más que separar “tipos” de validez. Esta reconceptualización indica que la evidencia para la validez se puede obtener a través de la examinación de cinco fuentes específicas: (1) la evidencia basada en el contenido, (2) la evidencia basada en los procesos de respuesta, (3) la evidencia basada en la estructura interna, (4) la evidencia basada en las relaciones hacia las otras variables, y (5) la evidencia basada en las consecuencias de las pruebas. Los estándares aclaran que la “validez” es un término global que define la suma de todos los esfuerzos hechos para demostrar la evidencia de la efectividad de las puntuaciones de las pruebas y sus interpretaciones sirviendo los propósitos previstos de las pruebas:

Estas fuentes de evidencia pueden iluminar los aspectos diferentes de validez, pero no pueden representar los distintos tipos de validez. La validez es un concepto unitario. Este es el grado por el cual toda evidencia acumulada respalda la interpretación prevista de las puntuaciones de las pruebas para los propósitos previstos (AERA, APA, & NCME, 1999, p.11).

Sin embargo la estructura de validación organizacional tripartita fue reemplazada con el concepto de examinar fuentes de evidencia, es importante señalar que, en la vasta mayoría de los manuales de prueba, los capítulos de validez prueba continúan estando divididos en secciones separadas que direccionan el contenido, criterio-relacionado, y la validez del constructo, incluso para las pruebas desarrolladas después de 1999. Goodwin (2002; Goodwin & Leech, 2003) ha notado el retraso significativo entre la emisión de los nuevos estándares y la adopción de estos estándares en los manuales de medida y los manuales de pruebas, y ha proporcionado un marco útil para traducir los conceptos de validez más antiguos en el nuevo marco basado en la evidencia. Las

secciones siguientes direccionan el tema de la validez de las pruebas de personalidad usando las cinco áreas de evidencia establecidos en el nuevo documento de estándares.

Evidencia basada en el contenido de la prueba

La evidencia del contenido de la prueba relaciona la razón para la construcción y la selección del ítem durante los procesos de desarrollo de la prueba. El desarrollo de una prueba de personalidad debe incluir la generación de un proyecto para guiar la selección del contenido y la organización estructural total de la prueba. Los argumentos lógicos, la revisión realizada por los expertos y, los análisis empíricos de los dominios de contenido y los ítems y, la data de prueba de ítem pueden ser presentados en el manual de la prueba para ofrecer evidencia de la adecuación de los esfuerzos de construcción de la prueba en esta etapa más temprana de desarrollo. Esta fuente incluye evidencia desde los análisis empíricos de la información del ítem para detectar posibles fuentes de preferencia (ejemplo, género, cultura, edad) que podría influir el uso válido de la prueba para los grupos específicos de evaluados. En muchas maneras, esta fuente coincide con varios aspectos de los antiguos conceptos de validez de contenido y de validez de constructo.

Esta fuente es donde los evaluadores tendrían expectativas para encontrar información acerca de la estructura subyacente de una prueba. Por ejemplo, ¿Sigue las divisiones del contenido de una prueba objetiva una categorización DSM-IV, o la estructura de alguna teoría de personalidad, o ambos? ¿Cuál es la razón subyacente para la constelación de las variables con puntuación del Rorschach? También lo oculto sería ¿En qué medida los ítems localizados en subescalas específicas parecen ser ajustados. Por ejemplo, ¿Qué evidencia se ofrece para apoyar los 10 ítems seleccionados para la colocación en una subescala pensada para evaluar Trastornos de Personalidad por Evitación? La evidencia que podría proveer para apoyar la afirmación que usa de una ficha de estímulo específico Rorschach no produce género ni parcialidad étnica en estimaciones de rendimiento en variables de puntuación específica Rorschach.

Evidencia basada en procesos de respuestas

La evidencia basada en los procesos de respuesta se refiere a la examinación de la evidencia que indica que los tipos de tareas o las respuestas de los ítems

requieren que los evaluados se encuentren en medios efectivos para la evaluación del contenido de la prueba. Esta fuente se ha conducido tradicionalmente en secciones de validez de constructo e incluye la razón para traducir el contenido en ítems o tareas, la justificación para los métodos usados para administrar los ítems o tareas y, una razón para el desarrollo de un sistema para los ítems de puntuación.

Las preguntas investigadas y presentadas en el caso de una prueba objetiva podría incluir lo siguiente: ¿Puede una prueba con papel y lápiz reflejar adecuadamente en una manera significativa percepciones, pensamientos, sentimientos y acciones del evaluado referente a uno o más rasgos de personalidad que la prueba está intentado medir? ¿Tiene la prueba una manera para demostrar que los evaluados están respondiendo a las preguntas de una manera honesta? La evidencia relacionada a la adecuación de sistemas de puntuación, una gran preocupación de las pruebas objetivas, sería también abordada en esta sección. Por ejemplo, ¿Qué evidencia se proporciona para apoyar los procesos de puntuación, criterios de puntuación y, el valor de la puntuación aplicada a las respuestas para los ítems de frases o incompletas o de dibujo de la figura humana? Para las pruebas objetivas, la evidencia podría ser ofrecida para justificar el uso de opciones de puntuación dicotómicos (verdad-falso o si-no) sobre las opciones de clasificación multipunto, o viceversa.

Evidencia basada en la estructura interna

La evidencia dirige la medida en que la estructura del contenido de la prueba corresponde a la estructura del constructo o constructos demostradas de manera empírica o teórico que la prueba es supuestamente medible. La evidencia que sostiene la estructura interna tradicionalmente ha sido el foco central de las discusiones de validez de constructo. Las aproximaciones empíricas para proporcionar la evidencia pueden incluir intercorrelaciones de los ítems, tales como aquellos conducidos para demostrar confiabilidad de consistencia interna, como esos análisis que demuestran en qué medida los ítems colocados en una escala o subescala producen patrones de respuesta similar. Los actuales análisis de factores es el método preferido de la mayoría de programadores de pruebas para proporcionar soporte empírico para las divisiones de escala y subescala propuestas dentro de una prueba (Nunnally & Bernstein, 1994). La popularidad de esta

técnica ha generado casi una aproximación monolítica para la evaluación de la evidencia de una estructura interna, tales que las pruebas que no muestran estructuras de factores sólidas son sospechosas en términos de adecuación para los propósitos clínicos. En realidad, una estructura de factor extremadamente separada algunas veces puede indicar que demasiada atención se ha pagado para ubicar los ítems en las escalas para producir la estructura de factor previsto y, muy poca atención ha sido pagada para la importancia del contenido de un ítem y los servicios clínicos de las puntuaciones que los factores relativamente puros proporcionan (Nunnally & Bernstein, 1994; Thorndike, 1997). La evidencia para apoyar la estructura interna podría dirigir la idoneidad de dividir un contenido de escala de prueba objetiva en subescalas esperadas para evaluar las características del trastorno de personalidad por Evitación, Borderline, Obesivo-Compulsivo, Esquizotípico y Paranoide en lugar de proporcionar solo una única Escala de Trastorno de Personalidad.

Evidencia basada en relaciones con otras variables

Esta fuente ofrece evidencia en la medida que los resultados de las pruebas se relacionan con los resultados obtenidos con otras pruebas u otros criterios tales como comportamientos observables y diagnósticos clínicos. Esta fuente coincide con las presentaciones tradicionales de validez relacionada al criterio, que incluye evidencias de validez actuales y predictivos y, evidencia de validez de constructo que normalmente involucra análisis discriminantes y convergentes de las relaciones de puntuación. Lo más importante, esta fuente proporciona soporte para los usos clínicos previstos de la prueba.

Los tipos más comunes de análisis de información tradicionalmente presentados en los manuales de prueba de personalidad para demostrar que la validez de constructo y el relacionado criterio involucran estudios correlacionados, especialmente comparaciones de pruebas con otras medidas pensadas para medirse en los mismos o diferentes rasgos o constructos de personalidad. Además, a menudo lo presentado son estudios que comparan las puntuaciones de pruebas para los grupos de muestras, por ejemplo, puntuaciones de pruebas para una muestra de niños o adolescentes no referidos comparados con las puntuaciones de pruebas de niños o adolescentes ya diagnosticados con uno o más trastornos de personalidad. Lo menos

presentado generalmente son estudios de diseño experimental, tales como estudios de intervención que usa puntuaciones de prueba como criterios de estatus pre y postintervención. Por ejemplo, si una conocida intervención tiene un gran triunfo con los pacientes adolescentes se utiliza y, la prueba se piensa que es una medida válida de cambio en las percepciones, pensamientos, sentimientos o acciones que son conocidas para cambiar de la exposición a la intervención, luego las puntuaciones de la prueba postintervención debe mostrar cambios en la dirección esperada cuando es comparada con las puntuaciones de la pruebas preintervención.

La evidencia relacionada a las consecuencias de la prueba

Esta fuente es planeada para dirigir las consecuencias que deriven del uso de la prueba. Estas consecuencias se podrían anticipar o no anticipar durante el comienzo de la prueba, y podría ser positiva o negativa por naturaleza. La especificación de esta fuente como una base para juzgar la validez de una prueba ha generado una gran controversia (Linn, 1997; Pophm, 1997; Shepard, 1997). A pesar de su naturaleza controversial, esta fuente se agregó en los estándares de 1999. Debido a su estatus relativamente nuevo como una fuente de evidencia para la validez de una prueba, ha recibido poca consideración en los capítulos de validez de manuales de pruebas de personalidad. Esta nueva área de consideración podría intentar dirigir preguntas tales como las siguientes: ¿Está el uso de las pruebas de la personalidad con los niños y adolescentes orientado a diagnosticar una buena o mala práctica profesional los trastornos de personalidad? ¿Ayuda a los niños el uso de los resultados de la prueba de personalidad obtener asistencia profesional y apoyo de la comunidad que podrían necesitar para que sus vidas puedan mejorar, o los marcan de manera permanente como anormal o reducen sus oportunidades de vida gratificante de importancia significativa? Si la prueba de personalidad demuestra evidencia de la habilidad para identificar o ayudar a identificar los objetivos del tratamiento ¿Son los objetivos derivados en esta manera aquellos que en realidad deben ser el centro de tratamiento y que por último beneficiarán al niño o adolescente? Si una prueba de personalidad demuestra sensibilidad para los cambios en las percepciones, pensamientos, sentimientos o acciones en el tiempo tal que se use para monitorear el avance en el tratamiento, ¿está la prueba realmente monitoreando los aspectos

de cambios importantes que deben estar ocurriendo si el niño o adolescente es realmente beneficiado en un periodo largo del tratamiento? Estas son preguntas difíciles que deben plantearse y discutirse en foros de objetivos abiertos.

La validación de la prueba y las cinco fuentes de evidencia

Aunque todas estas fuentes se pueden usar para desarrollar un argumento para apoyar la validez de una prueba, no es necesario que cada prueba proporcione información para todas las cinco de estas fuentes. En algunos casos, una fuente puede ser mucho más importante para establecer la validez de una prueba que otras. El juzgar la validez de una prueba puede ser un problema complicado, el cual puede ser más fácil para mantener el foco en el aspecto de validez más importante: demostrar la evidencia que la prueba puede ser usada efectivamente y con buenos beneficios con la finalidad por la que está siendo usada.

Es fundamental apreciar que algunos usos potenciales de una prueba podría no anticiparse por los creadores de las pruebas. En estos casos, los estudios externos de los procesos de desarrollo de las pruebas deben ser conducidos para proporcionar evidencia que apoye nuevos usos de la prueba. El hecho que este nuevo uso no se manifieste como uno de los usos originales de la prueba es menos importante que la evidencia substancial proporcionada para apoyar el nuevo y no anticipado uso de la prueba. Teniendo más de 20 años de experiencia en el campo del desarrollo de la prueba, el primer autor a menudo se plantea una sola interrogante que pertenece a varias pruebas: “¿Es la prueba x una prueba válida?” la respuesta proporcionada siempre es la misma: “eso depende. ¿Cuál es la finalidad del uso de la prueba x?”. La validez no se puede establecer independientemente de los usos previstos de una prueba.

En los casos de las pruebas de personalidad usadas con niños y adolescentes sospechosos de demostrar los trastornos de personalidad, los usos más probables de prueba sería el de ayudar con diagnósticos de problemas psicológicos, para el plan del tratamiento, y de monitorear los efectos del tratamiento. Para que se juzgue válido con fines de diagnóstico de los trastornos de la personalidad, la prueba necesitaría mostrar evidencia que la interpretación del puntaje de la prueba efectivamente identifica los trastornos de personalidad

o proporciona una mejora gradual en proporcionar diagnósticos de trastornos de personalidad sobre el uso de otras técnicas empleadas sin una prueba. Relacionado con el planeamiento del tratamiento, los creadores de la prueba necesitarían proporcionar evidencia que la interpretación del rendimiento de la prueba se puede usar efectivamente para planear el tratamiento de una manera que es igual o que mejora el planeamiento sobre las técnicas que no incluyen el uso de los resultados de las pruebas. Con la finalidad de monitorear los efectos del tratamiento, los creadores de las pruebas necesitarían demostrar que la prueba es sensible a cambios en las acciones, sentimientos, pensamientos, percepciones de los niños o adolescentes considerados ser representativos de los rasgos de la personalidad que la prueba pretende medir. Finalmente, el tema de evidencia de las consecuencias del uso de la prueba necesitaría ser dirigido en el manual de prueba y discutida abiertamente en la literatura profesional. El resultado de tales discusiones debe ser evidencia de la forma final mencionada en los Estándares: de la evidencia basada en las consecuencias de la prueba. Con optimismo, tal evidencia apoyará la afirmación de que la prueba es una herramienta útil que puede ser puesta para un buen uso por el beneficio de los niños y adolescentes con quienes se podría usar.

VISIÓN GENERAL DE LA EVALUACIÓN DE LOS TRASTORNOS DE LA PERSONALIDAD DEL NIÑO Y ADOLESCENTE

Este capítulo se ha centrado en las características necesarias para una prueba estandarizada de personalidad o trastornos de personalidad para ser considerada una herramienta de evaluación confiable y válida. Sin embargo, como se señaló anteriormente, las pruebas son solo una de los varios métodos de evaluaciones que pueden y deben ser usados cuando intentan identificar niños y adolescentes que cumplen con los criterios de trastornos de personalidad. Consistente con las mejores prácticas generales para la evaluación de los problemas de comportamiento mostrados por los niños, la evaluación relacionada específicamente con los diagnósticos y el tratamiento de los trastornos de personalidad en los niños deben incluir un enfoque multifuente y multimétodo que incluya entrevistas con el niño o adolescente y padres y cuidadores, la administración de uno o más medidas de autoinforme adecuados al desarrollo, y la colección de la información colateral desde tanta muchas fuentes

como sea posible. Esta sección discute el rol de cada uno de estos métodos de evaluación en el proceso de evaluación global.

La colección de la información colateral y la revisión de los registros

Reunir información desde las fuentes colaterales es la piedra angular de la evaluación de los trastornos de personalidad en la etapa de la niñez y la adolescencia. La aportación de los padres o de los principales cuidadores del niño o adolescente es considerado esencial para un diagnóstico válido. Los padres normalmente tienen el más elevado grado de familiaridad con el pensamiento del niño o adolescente y los patrones de comportamiento a través de los más largos periodos de tiempo, al menos en el caso de las familias intactas y estables. En los casos donde la estructura o función familiar se ha desestabilizado significativamente (divorcio, muerte de un padre, abandono, adopción, acuerdos de adopción, etc.), la aportación de los principales cuidadores es crucial, aunque la precisión y la confiabilidad de la información proporcionada en estas circunstancias podrían ser menores a la deseada. Siempre que sea posible, es prudente reunir información de más de un padre o tutor. La data proveniente de los miembros de la familia múltiple incrementa la probabilidad de que la aportación de la familia es confiable y válida. Para ayudar en la determinación de si el pensamiento trastornado y los patrones de comportamiento son persistentes a través de los escenarios y situaciones, a menudo se requiere de la aportación de los profesores del salón y el otro personal de la escuela quienes interactúan de manera frecuente con el niño y el adolescente. Las políticas y prácticas educativas también posibilitan la colección más efectiva y revisión de los registros históricos que es el caso con los pacientes adultos. Estos registros pueden ser extremadamente útiles para determinar en qué medida el pensamiento trastornado y los patrones de comportamiento representan problemas de mucho tiempo atrás o solamente acontecimientos transitorios.

Una técnica elemental para la colección de información de las fuentes colaterales para los problemas de comportamiento en la etapa de la niñez y la adolescencia es la terminación de las escalas de clasificación del comportamiento dada por los padres u otros cuidadores y profesores. Las dos escalas más usadas para este propósito son, El Sistema de

Evaluación del Comportamiento por el Niño- Segunda Edición (BASC-II; C.R.Reynolds&Kamphaus, 2004) y el Sistema de Evaluación Basado Empíricamente de Achenbach (Achenbach & Rescorla, 2004) Child Behavior Checklist (Lista de Verificación de Comportamiento para el Niño, por sus siglas en inglés CBCL) y Teacher Report Form (Formulario de Reporte del Profesor, por sus siglas en inglés TRF). Estas escalas han demostrado su utilidad en la evaluación de los problemas del comportamiento y lo emocional, incluyendo trastornos psiquiátricos tales como la depresión, la ansiedad, trastornos de conducta, aunque en un principio no se diseñaron para proveer información de los calificadores acerca del suceso de los pensamientos y comportamientos asociados con los trastornos de personalidad específico como se describe en el DSM.IV.

A pesar del extenso uso de estas medidas para la evaluación de los problemas emocionales y de comportamiento de Eje I su utilidad en la evaluación de la psicopatología de la personalidad es prácticamente desconocida. Una revisión de la CBCL sugiere que aunque los ítems se seleccionaron para evaluar los problemas y los síndromes del Eje I, muchos ítems pueden ser útiles en la identificación de rasgos de personalidad maladaptativos y sufrimiento específico de niños y adolescentes. Por ejemplo, los ítems tales como “el miedo a que él o ella pueda pensar o hacer algo malo”, “tímido”, y “acomplejado o fácilmente avergonzado” pueden ser indicadores potenciales de las características de personalidad evitante o timidez extrema. Adicionalmente, los ítems tales como “los actos muy jóvenes para su edad” y “aferrarse a los adultos o ser muy dependiente” pueden ser indicadores de rasgos de personalidad dependiente. “Los que jumbroso” o “los que tienen comportamientos extraños”, “los que prefieren estar solo que con otras personas”, “tienen ideas extrañas”, y “son introvertidos, es decir no se involucran con los demás” pueden ser indicadores de rasgos de personalidad esquizoide.

Aunque puede haber correspondencia entre los numerosos ítems en las mediciones de Achenbach y otros problemas de personalidad y escalas de clasificación del niño y adolescente, no está del todo claro si los fenómenos del Eje II pueden ser identificados por el CBCL y por el TRF. Las búsquedas de la literatura para la investigación que usa el CBCL, TRF, o BASC en la evaluación de los trastornos de personalidad revelaron

solo un estudio. Bradely, Conklin, y Westen (2005) usó el CBCL, como lo completado por los especialistas, para ayudar a validar de manera empírica a los subgrupos de mujeres adolescentes que cumplen con los criterios de trastorno de personalidad (BPD). Ellos son derivados y etiquetados en 4 grupos de chicas con BPD basado en su metodología Q: la interiorización de funcionamiento alto, la interiorización depresiva, histriónico, y exteriorización de ira. De apoyo a sus hipótesis, Bradley et al. Encontró que los análisis de validación adicional mostraron que el grupo de interiorización depresiva tuvo más altos puntajes en la escala de interiorización del CBCL que los otros grupos, mientras que el grupo de exteriorización de ira tuvo los más altos puntajes en la escala de exteriorización. Consecuentemente, los resultados de la CBCL proporcionó algunas validaciones para el método de establecimiento de subgrupos de chicas con BPD de Bradley et al. Esto sugiere que las escalas de banda ancha de CBCL pueden ser dimensiones importantes en el cual se inicia para identificar más dimensiones de personalidad de psicopatología más duraderos. Sin embargo, cabe mencionar que Bradley et al. usaron las clasificaciones de los especialistas más que las clasificaciones de los padres. Por ello, la evidencia empírica hasta ahora es débil en lo que respecta al uso de escalas de clasificaciones populares del profesor y del padre de psicopatología del niño. Claramente, se necesita más investigaciones en esta área y debe ser una gran prioridad para los investigadores que estudian los problemas del Eje II en la juventud.

La escala psicopatológica del adolescente (por sus siglas en inglés, APS) (W.M Reynolds, 1998) es la única medida de autoinforme que se puede usar con adolescentes que, en parte, se diseñó específicamente para dirigir los trastornos de personalidad como se definió en el DSM-IV. La escala psicopatológica del adolescente proporciona escalas de contenido numerosos que evalúan los síntomas de muchos trastornos psicológicos descritos en DSM-IV, que incluyen escalas específicas que evalúan la severidad de síntomas relacionados a los cinco de los trastornos de personalidad del DSM-IV (evitante, obsesivo-compulsivo, borderline, esquizotípico, y paranoide). La información de validez reportada en el manual técnico de la APS parece acertada y, el instrumento muestra una gran promesa para los usos clínicos y de investigación en relación a los 5 trastornos de personalidad que éste evalúa. Resulta importante

señalar que aunque la metodología del autoinforme de la prueba objetiva parece ser una técnica válida para obtener información acerca del funcionamiento psicológico de adolescentes y los niños mayores (W.M Reynolds, 1993), la aplicación específica del método para el diagnóstico de los trastornos de personalidad no está aún muy bien establecida. Los especialistas deben calibrar cuidadosamente la confiabilidad de los autoinformes proporcionados por los entrevistados adolescentes quienes realmente muestran trastornos de personalidad certeros.

TÉCNICAS DE ENTREVISTA

Un número de investigadores han sugerido que las entrevistas clínicas son el método principal de evaluación de los trastornos de personalidad en la práctica clínica general (Watkins, Campbell, Nieberding, & Hallmark, 1995; Westen, 1997; Widiger & Samuel, 2005). Aunque las entrevistas no estructuradas pueden ser útiles en la obtención de información acerca de los pacientes de una manera no amenazante, se han reportado varias desventajas con su aplicación. Los problemas con las entrevistas no estructuradas comprenden errores al incorporar la discusión de todos los criterios de diagnóstico necesario (Blashfield & Herkov, 1996; Zimmerman & Mattia, 1999), exagerando similitudes entre el estímulo no idéntico a través de la omisión de la variabilidad dentro del grupo, considerando la evidencia no confirmada, centrándose en los ejemplos estereotípicos de la categoría (Cantor & Genero, 1986), diagnosticando de manera jerárquica y, equivocándose en la evaluación de los síntomas adicionales una vez que un trastorno particular se ha identificado (Alder, Draker, & Teague, 1990; Herkov & Blashfield, 1995). Los intereses idiosincrásicos de los especialistas también pueden influir en los diagnósticos asignados al paciente (Gunderson, 1992; Mellsop, Varghese, Joshua, & Hicks, 1982). Además de los inconvenientes metodológicos, los pocos estudios que se han llevado a cabo para evaluar la confiabilidad de las entrevistas no estructuradas para el diagnóstico del trastorno de la personalidad no han sido fomentados (Blashfield & Herkov, 1996; Zimmerman & Mattia, 1999).

Las entrevistas semiestructuradas se han desarrollado para dirigir los defectos de técnicas de entrevistas poco estructuradas y se han investigado de manera más amplia en términos de su aplicabilidad

para el diagnóstico de los trastornos de personalidad. El diseño de las entrevistas semiestructuradas asegura que una consideración comprensiva y sistemática de todos los trastornos de personalidad ocurra (Widiger & Samuel, 2005). Aunque normalmente se asume que los entrevistados responderán de una manera confiable y abierta a las preguntas de la entrevista, esta suposición de apertura puede ser injustificada en el caso de los niños y adolescentes que exhiben ciertos trastornos de personalidad. Además, algunos investigadores han encontrado que la confiabilidad y la validez de las entrevistas del niño puede estar faltando seriamente en muchos casos, y ser dudosa en cierto grado, incluso en las mejores condiciones (Myers & Winters, 2002). Como se señaló anteriormente, la confiabilidad de un procedimiento de evaluación no es simplemente una función del desarrollo de metodología de la evaluación, sino también es una función de la interacción entre la metodología de evaluación y el individuo evaluado. Como en los autoinformantes, los niños y los adolescentes son mucho menos confiables que los adultos, y cuanto más joven sea el niño, la información obtenida es menos confiable (McCloskey, 1990). Aunque es importante entrevistar a un niño o adolescente sospechoso de tener un trastorno de personalidad, los especialistas deben tener cuidado en obtener información de las múltiples fuentes colaterales para verificar la exactitud o corroborar la credibilidad de las oraciones hechas por el niño o adolescente.

Una entrevista menos estructurada para los niños y adolescentes es la Entrevista de Evaluación de la Personalidad (PAI, por sus siglas en inglés; Selzer, Kernberg, Fibel, Cheruliez & Mortati, 1987). El PAI es una entrevista de 45 minutos con orientación psicoanalítica diseñada para evaluar las autorepresentaciones y las representaciones objetivas, el afecto, el pensamiento, la capacidad de reflexión y la empatía de los niños y adolescentes con el entrevistador (Kernberg, Weiner & Bardenstein, 2000). Esta medida se adaptó de la entrevista estructural de Kernberg's (1988) diseñada para evaluar las constelaciones estructurales básicas de la personalidad - neurótica, borderline y psicótica- y es analizada por Kernberg et al. La información reunida del PAI es usada así para ayudar a realizar determinaciones acerca de la estructura de personalidad y los posibles diagnósticos Eje II.

Sin embargo, esta técnica de entrevista aparece limitada en varias maneras importantes. Primero, aunque la entrevista se deriva de forma teórica, sus propiedades empíricas no son conocidas. No está disponible ninguna investigación para evaluar ni la validez ni la confiabilidad de esta medida. Consecuentemente, no queda exactamente claro qué constructos y qué tan confiable evalúa la entrevista. Segundo, la operacionalización de los constructos teóricos del PAI, se cree que la medida parece incompleta. Los constructos tales como las representaciones objetivas y las autorepresentaciones son ideas altamente abstractas acerca de las operaciones mentales pero que no están muy bien articuladas dentro del fenómeno medible y observable. Tercero, la falta de estandarización hace esta medida particularmente vulnerable a una variedad de los problemas metodológicos descritos previamente, incluyendo similitudes de exageración por la variabilidad de las omisiones dentro del grupo, discutiendo evidencia disconforme, enfocándose en los ejemplos estereotípicos de categorías y, los enfoques potenciales de los intereses idiosincráticos del especialista. Cuarto, no está claro cómo la información recogida de la medida esquematiza los constructos de trastorno de personalidad, definido por el DSM-IV. Kernberg et al. (2000) discute los resultados del PAI en términos de tipos de "organización" de personalidad pero no articula cómo estos se tradujeron en diagnósticos de Eje II.

Un enfoque más confiable y válido para la entrevista puede ser el uso de las entrevistas estructuradas y semiestructuradas. Actualmente, las únicas entrevistas directas estructuradas y semiestructuradas desarrolladas para el uso con los niños y adolescentes como informantes principales son el programa de entrevista diagnóstico diagnostic interview schedule for children, versión 2.3 (Shaffer et al., 1996), y la entrevista diagnóstica con los niños y adolescentes Diagnostic Interview for Children and Adolescents (Reich, 1998), y estos no abordan directamente temas relacionados específicamente con los trastornos de personalidad de la etapa de la niñez. La entrevista clínica estructurada para los trastornos de personalidad (SCID-II, por sus siglas en inglés) DSM-IV (Primero, Gibbon, Spitzer, Williams, & Benjamin, 1997), usada a menudo en la evaluación de adultos, ha demostrado algunas promesas en la evaluación de los trastornos de personalidad del adolescente. En su investigación

que comprende la evidencia basada de los trastornos de personalidad, Widiger y Samuel (2005) señalan que el SCID-II estipula que los especialistas necesitan verificar la presencia de los criterios de diagnóstico en un periodo de tiempo de 5 años. Cuando se evalúa a un adulto muy joven usando el SCID-II, todo o parte del periodo de 5 años antes de la entrevista podría abarcar los años de adolescencia. En un esfuerzo para examinar la asociación entre la psicopatología de la niñez y los trastornos de personalidad del adulto joven, se utilizó el SCID-II de 8 a 10 años después de la primera evaluación a los participantes de estudios (Kasen, Cohen, Skodol, Johnson, & Brook, 1999). Al comienzo del estudio Children in the Community (CIC), un estudio longitudinal de los factores de riesgo para la psicopatología del niño se inició en el año 1970, los participantes tenían de 1 a 10 años de edad. La primera sesión siguiente tuvo lugar 8 años después del comienzo del estudio. El SCID-II se usó durante la primera y todas las sesiones posteriores. Los ítems extraídos del SCID-II se usaron para ayudar a hacer los diagnósticos DSM-III-R de los trastornos de personalidad Evitante, Borderline, Dependiente, Histriónico, Narcisista, Obsesivo-Compulsivo, Paranoide, Pasivo-Agresivo, Esquizoide y Esquizotípico en el joven. La data SCID-II se combinó con la data de la entrevista y la pregunta recogida de los niños y sus madres para los diagnósticos de Eje II. Por ello, no se usó solo el SCID-II.

Sin embargo, los procedimientos de los diagnósticos en este estudio, incluyendo el SCID II, aparecieron para demostrar la validez, como se evidenció por varios estudios de muestra siguientes. Bernstein et al. (1993) encontraron que los diagnósticos de trastornos de personalidad entre los niños y adolescentes fueron asociados con los indicadores múltiples de angustia y discapacidad, que incluyen relaciones románticas y de parejas de calidad deficiente, mayor rendimiento escolar y de trabajo y problemas de comportamiento y mayor probabilidad de contacto con la policía, y una gran psicopatología de Eje I. Adicionalmente, Cohen et al. (2005) mostraron que los trastornos de personalidad Cluster A excéntrico/raro diagnosticado en la adolescencia fueron posteriormente predictivos de ajustes deficientes durante la transición a la etapa adulta, como lo indicado por menos años de formación o educación postsecundaria y la paternidad temprana.

Por consiguiente, la evidencia disponible sugiere que el SCID-II puede ser una entrevista válida e importante, usada en los diagnósticos de fenómeno Eje II en niños y adolescentes. Sin embargo, la evidencia también sugiere que puedan existir algunas limitaciones para el SCID-II. Lo más importante es la falta aparente de estabilidad del diagnóstico en la muestra CIC. Como lo señalaron Bernstein et al. (1993) menos de la mitad de los sujetos que recibieron el diagnóstico Eje II en tiempo 2 no recibieron ningún diagnóstico Eje II en tiempo 3, solo 2 años más tarde. Asimismo, la prevalencia de los trastornos de personalidad moderada y grave alcanzó su punto máximo en los 12 años para los niños y la de 13 años para las niñas. A diferencia de los incrementos teóricos en los trastornos de personalidad como los procedimientos de desarrollo, los diagnósticos descendieron de su punto más alto en la adolescencia temprana a través de los años de adolescencia. En su conjunto, estos hallazgos indican deficiencias del SCID-II, u otras medidas, como herramientas de diagnóstico. Además, parece que los diagnósticos hechos al menos en parte por el SCID-II pueden mostrar falta de estabilidad a corto plazo así como declinar cursos durante la adolescencia.

LAS PRUEBAS DE PERSONALIDAD

Las medidas de autoinforme objetiva y los métodos proyectivos se pueden aplicar para reunir información directamente de los niños y adolescentes. Las dos medidas de autoinforme adulta más utilizada, el Inventario Multifásico de Personalidad de Minnesota II (Minnesota Multiphasic Personality Inventory II, MMPI, por sus siglas en inglés) y el Inventario Clínico Multiaxial de Millon III (Millon Clinical Multiaxial Inventory III), también tienen que corresponder a las versiones del adolescente: el Inventario Multifásico de Personalidad de Minnesota II – Adolescente (MMPI-A, por sus siglas en inglés; Butcher et al., 1992) y el Inventario Clínico Adolescente de Millon (MACI por sus siglas en inglés; Millon, Millon, Davis, & Grossman, 1997). Aunque el MMPI-A no contiene escalas que corresponden directamente a las clasificaciones de los trastornos de personalidad DSM-IV, sí proporciona las 5 escalas psicopatológicas de la personalidad (agresividad, psicoticismo, falta de control (disconstraint), emocionalidad negativa/neuroticismo, introversión/escasa emocionalidad positiva) y otras escalas de contenido que direccionan las percepciones, pensamientos, sentimientos y

comportamientos característicos de los adolescentes que muestran trastornos de personalidad. Asimismo, el MACI no tiene escalas específicas que correspondan a las clasificaciones de trastorno de personalidad DSM-IV, en cambio proporciona configuraciones de escala de contenido múltiple y varios conjuntos de escalas de contenido que evalúan los patrones de personalidad, tal como la tendencia borderline, la autohumillación, lo conflictivo, lo conformista, lo egoísta, lo rebelde, lo vehemente. Cada una de estas escalas de contenido comprende subescalas múltiples; por ejemplo, la escala de tendencia borderline incluye tres subescalas: Temperamentalmente emotivo, cambios cognitivos, autoimagen incierta. Tal y como sus nombres implican, la MMPI-A y la MACI evalúan solo las características de la personalidad de los adolescentes de la edad de 12 a 18 años. Aunque el MMPI no tiene una escala que se designe para evaluar a los niños preadolescentes, el MACI sí tiene un equivalente, el inventario clínico preadolescente de Millon (M-PACI, Tringone, Millon, & Kamp, 2007). El M-PACI se diseñó para utilizarlo con niños de 9 a 12 años de edad, sin embargo las escalas de contenido son de carácter general y no están atadas directamente con las clasificaciones de trastorno de personalidad. Las dos escalas usadas con mayor frecuencia para evaluar los problemas de comportamiento general en los niños – el BASC-II y el CBCL- tienen versiones de autoinforme del niño y adolescente, pero las estructuras de esas medidas de autoinforme corresponden estrechamente a aquellas versiones del padre y del profesor y por ello no contienen escalas que se alineen con ninguna clasificaciones de trastorno de personalidad DSM-IV.

Varias técnicas proyectivas se han usado para evaluar múltiples aspectos de la personalidad y la psicopatología del niño y adolescente (Knoff, 1986). Estos incluyen los test de mancha de tinta de Rorschach, el Test de Apercepción Temática (Murray, 1943), pruebas gráfica-proyectivas tales como la prueba del dibujo de una persona (Machover, 1949), Prueba de frases incompletas (Westenberg, Treffers, & Drewes, 1998). Sin embargo, la confiabilidad y validez de las medidas proyectivas para la evaluación de la psicopatología de la personalidad siguen siendo cuestionables. Asimismo, como es el caso con las medidas de autoinforme, el uso de las técnicas proyectivas en el diagnóstico de los trastornos de personalidad en los niños y adolescentes aún no se ha establecido muy bien en la literatura profesional.

LA METODOLOGÍA Q-SORT

El trabajo de Westen en el diagnóstico y evaluación de la personalidad de los problemas Eje II en las posiciones de los adolescentes contrasta la literatura y los métodos limitados que rodean otras medidas de personalidad numerosas. Shedler y Westen (1998) han intentado desarrollar un enfoque alternativo para evaluar los trastornos de Eje II. Su método, el Procedimiento de Evaluación de Shedler-Westen (SWAP, por sus siglas en inglés), ha experimentado tres revisiones. La primera versión, el SWAP-167, fue una versión del ítem 167, el cual fue sometido a prueba en los trastornos del Eje II Cluster B (Shedler & Westen, 1998). El SWAP-200 se reguló en una muestra de 530 pacientes trastornados de personalidad. La versión actual, el SWAP-II, se regula en una muestra de 1 201 pacientes con patología de personalidad de leve a severa. Se desarrolló también una versión de adolescente del SWAP-200, el SWAP-200-A y, se reguló en 294 adolescentes que exhibieron un rango de patología de personalidad (Westen, Shedler, Durrett, Glass & Martens, 2003). La versión actual, la SWAP-II-A se está regulando en 950 adolescentes con patología de personalidad de leve a severa (Westen, Shedler & Bradley, 2005).

El SWAP utiliza la metodología Q-sort de evaluación de personalidad. Esto consta de 200 ítems derivados de varias fuentes, que incluyen los criterios del diagnóstico de trastorno de personalidad del texto y los anexos del DSM-III-R y DSM-IV; ítems seleccionados del Eje I que podría reflejar patología de personalidad; un número importante de descriptores de personalidad derivados de la literatura de la personalidad y los trastornos de personalidad escritos en los últimos 50 años; investigación en los procesos de adaptación y defensivos; investigación en la patología interpersonal de los pacientes con trastornos de personalidad; investigación en los rasgos y desarrollo de personalidad normal; y entrevistas extensas piloto. Los ítems desarrollados por el SWAP-II-A también tomó en consideración la investigación de la psicopatología, personalidad y desarrollo en el adolescente. Para proporcionar un perfil completo de las fortalezas de un paciente además de la patología, muchos de los ítems se diseñan para evaluar aspectos de la patología de la personalidad que no son lo suficientemente severas para garantizar un diagnóstico del Eje II, así como los aspectos de funcionamiento saludable (Westen et al.,

2005). Un especialista o entrevistador que conoce a los pacientes debe clasificar muy bien las tarjetas en categorías basadas en el grado en el cual describen de manera precisa al paciente. El SWAP-II requiere de especialistas que clasifiquen las 200 oraciones en 8 categorías, que van desde el 0 (irrelevante o inaplicable) hasta el 7 (altamente descriptivo). Las oraciones intentan proporcionar un vocabulario estándar para ayudar a los especialistas a expresar sus observaciones e inferencias de una manera que se pueda cuantificar, comparado con las descripciones de otros especialistas, y analizado estadísticamente. Los autores notan que hay dos maneras principales que el procedimiento SWAP difiere en los criterios del Eje II. La primera es que los ítems se desarrollaron para operacionalizar los constructos psicológicos, tales como los procesos defensivos, que son excluidos de la DSM-IV debido a la preocupación que esos procesos no puedan ser medidos de manera confiable. Segunda, el rango de los ítems se expande para incluir aspectos de funcionamiento que podría ser clínicamente importante pero que el Eje II no lo dirige, tales como los propósitos, conflictos entre los propósitos, estrategia de regulación efectiva, aspectos leves de patología de personalidad, y áreas de funcionamiento adaptativo positivo (Westen & Shedler, 1999). Cuando se usa la metodología SWAP-II es importante señalar que los especialistas no deciden cuantos ítems colocar en cada categoría; en cambio, los ítems se deben organizar en distribuciones fijas con un número específico de oraciones en cada categoría. Este procedimiento se usó para la ventaja de las pruebas psicométricas, tales como la minimización del error de medida (Westen et al., 2005). La distribución fija se modeló después de un medio derecho de una distribución de curva normal y de forma de campana (Westen & Shedler, 1999). Por ello, muchos ítems recibirán bajos valores (0,1 y 2), sin embargo unos pocos recibirán valores de 7. Existe una versión electrónica de la metodología Q SWAP-II que puede ser utilizada.

El SWAP-II (para adultos) y el SWAP-II-A (para adolescentes) recientemente se usaron en dos extensos estudios fundados por el Instituto Nacional de Salud Mental que se condujeron como parte de la revisión del Eje II (Westen et al., 2005); la data del análisis se está llevando a cabo actualmente. El primer estudio utilizó un extenso número de especialistas de nivel doctoral a lo largo de América del Norte para proporcionar datos cuantificados en un paciente a su

cuidado. La data recogida se analizará para identificar las mayores características así como las características más distintivas de cada trastorno. Los análisis de factor Q y los análisis de clase latente serán usados para identificar de manera empírica grupos de diagnósticos independientemente de los criterios actuales del DSM-IV. La data se puede analizar para identificar la estructura de rasgo subyacente para cada trastorno de personalidad mientras que retiene los constructos de trastornos de personalidad. Finalmente, la data se puede usar para identificar la estructura del rasgo de la patología de personalidad independientemente de los diagnósticos actuales.

En otro reciente estudio por Shedler, Westen, y Bradley (2006), una muestra nacional aleatoria de 291 psiquiatras y psicólogos clínicos describen un paciente seleccionado aleatoriamente en su cuidado. La data del diagnóstico proporcionado por el especialista se usó para generar los diagnósticos DSM-IV categórico y dimensional (número de síntomas presentes por el trastorno). Los especialistas también utilizaron uno de los dos prototipos de sistemas de correspondencias para proporcionar un diagnóstico para el paciente seleccionado. Los análisis de los resultados encontraron que el diagnóstico del prototipo produjo comorbilidad reducida correspondiente al diagnóstico DSM-IV; produjo estimaciones similares de validez en la predicción de variables de criterios, tales como el funcionamiento adaptativo, respuesta de tratamiento, y la etiología; y fue mejor que el diagnóstico DSM-IV en el rango de utilidad clínica y facilidad de manejo. Basados en sus resultados, los autores concluyeron que agregar un prototipo más de salud de personalidad incrementa la precisión de predicción. El estudio encontró la evidencia de que un simple prototipo de procedimiento combinado proporciona una alternativa viable para mejorar los diagnósticos de trastornos de personalidad en la práctica clínica. Shedler et al. concluyó que un prototipo de diagnóstico es fácil de manejar, que minimiza la comorbilidad artefactual, que es compatible de que ocurra naturalmente procesos cognitivos, y que trasfiere de inmediato en ambos diagnósticos categórico y dimensional. Aunque el SWAP-II-A se encuentra aún en la etapa temprana de desarrollo, muestra un gran compromiso para aumentar y remodelar significativamente prácticas de evaluaciones actuales relacionadas con los trastornos de personalidad.

CONCLUSIONES

La evaluación de los niños y adolescentes con el propósito de establecer un diagnóstico de trastorno de personalidad es una práctica relativamente nueva, a pesar de investigación amplia y creencias clínicas de toda la vida correspondiente a la importancia de los atributos de personalidad de los niños para la psicopatología. Por lo demás, existe un número de estudios publicados que sugieren problemas psicológicos de niños y adolescentes maladaptativos, inflexibles y duraderos que se pueden evaluar y se encuentran en un área legítima de estudio. Un número de estudios sugiere que los trastornos de personalidad borderline, esquizoide y evitante, además de otros, se pueden evaluar durante la etapa de niñez y la adolescencia y, como se espera, se relacionan con el desarrollo de calidad muy pobre.

A pesar de estos éxitos, el estado actual de la literatura en la evaluación del fenómeno del Eje II en los niños puede describirse mejor como carencia en la mayor parte. Las prácticas de evaluación actuales que envuelven pruebas de personalidad son extensiones decrecientes de las técnicas usadas con los adultos. Como lo señala Bradley et al. (2005) al menos para el Trastorno de Personalidad Borderline, los estudios numerosos han usado criterios de adulto para diagnosticar a los adolescentes con éxito aparente. En las evaluaciones basadas en los criterios de adulto se han identificado adecuadamente adolescentes con psicopatología importante así como las discapacidades asimismo han sido validadas en contra de otros métodos de evaluación. Sin embargo, la pregunta aún permanece si esos son los criterios óptimos para los diagnósticos. Este no es un problema limitado para investigar el Trastorno de Personalidad Borderline. La evidencia para los criterios de diagnóstico óptimo para prácticamente todos los trastornos de Eje II, e incluso en condiciones del Eje I tales como el trastorno bipolar y la depresión mayor, es confusa.

Quizás la literatura existente en este sentido es un buen comienzo. Sin embargo parece que una de las metas más importantes para la futura investigación en la psicopatología de la personalidad del niño es perfeccionar y establecer criterios de diagnóstico fundamental.

Una limitación adicional es la ausencia de una investigación que establezca medidas de estándares de psicopatología Eje I, tales como el CBC y TRF, como medidas de problemas Eje II. Sin embargo, muchos ítems incluidos en estas medidas para identificar problemas mayores del Eje I también se puede usar para identificar rasgos maladaptativos y persistentes. Una revisión de las escalas de Achenbach revela muchos ítems que pueden utilizar un rango de trastornos de personalidad que comprende el Esquizoide, Borderline, y Evitante. Las escalas más populares tales como el CBCL, TRF y BASC se pueden utilizar para la evaluación de los problemas de personalidad. Sin embargo, la literatura es bastante limitada en demostrar el valor empírico en este enfoque. Prácticamente no se ha identificado ningún estudio que haya probado la confiabilidad y validez de las escalas de rango Eje I mayor en las evaluaciones de los trastornos de personalidad en los niños y adolescentes. Además, la literatura en otras mediciones, tales como las pruebas de la personalidad, las pruebas proyectivas, y las entrevistas es también bastante restringida. Por consiguiente, una meta mayor para el campo debe ser un examen riguroso de la existencia de mediciones para determinar la validez y confiabilidad en la evaluación.

Se necesita también una mayor atención. La mayoría de literatura existente se centra en los adolescentes. Consecuentemente, hay una gama más amplia de medidas y una evidencia base más desarrollada para la evaluación de los problemas de personalidad entre adolescentes. Algunas de estas literaturas han sido innovadas, particularmente el trabajo de Westen y sus colegas, y el trabajo inicial de validez de ambas medidas y criterios de diagnóstico han dado resultados prometedores. No es de extrañar que se haya realizado más trabajo en muestras con adolescentes dándose la tendencia entre investigadores para expandir los métodos de literatura adulta a los grupos de edad joven. Sin embargo, la atención en los adolescentes ha resultado en un alcance incluso más limitado de niños. Aunque algunas de las literaturas sugiere que la evaluación de los niños es posible (Kasen et al., 1999), se necesita más investigación.

REFERENCIAS

- Adler, D. A., Drake, R. E., & Teague, G. B. (1990). Clinicians' practices in personality assessment: Does gender influence the use of DSM-III Axis II? *Comprehensive Psychiatry*, 31, 125-133.
- American Educational Research Association (AERA), American Psychological Association (APA), y National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Asociación Psiquiátrica Americana. (2000). *Manual diagnóstico y estadístico de los trastornos mentales* (4ª ed. texto revisado). Washington, DC: Autor.
- Blashfield, R. K., & Herkov, M. J. (1996). Investigating clinician adherence to diagnosis by criteria: A replication of Morey and Ochoa (1989). *Journal of Personality Disorders*, 10, 219-228.
- Bradley R., Conklin, C.Z. & Westen D. (2005). The borderline personality diagnosis in adolescents: gender differences and subtypes. *Journal of Child Psychology and Psychiatry*, 46(9), 1006-1019.
- Cantor, N., & Genero, N. (1986). Psychiatric diagnosis and natural categorization: A close analogy. In T. Millon & G. Klerman (Eds.), *Contemporary directions in psychopathology* (pp. 233-256). New York: Guilford Press.
- Gunderson, J. G. (1992). Diagnostic controversies. In A. Tasman & M. B. Riba (Eds.), *Review of psychiatry* (Vol. 11, pp. 9-24). Washington, DC: American Psychiatric Press.
- Herkov, M. J., & Blashfield, R. K. (1995). Clinicians' diagnoses of personality disorder: Evidence of a hierarchical structure. *Journal of Personality Assessment*, 65, 313-321.
- Hilsenroth, M. & Segal, D. (Eds.) (2004). Personality assessment. In M. Hersen (Editor-in-Chief), *Comprehensive handbook of psychological assessment* (Vol. 2). Hoben, NY: John Wiley & Sons.
- Kasen, S., Cohen, P., Skodol, A., Johnson, J.G., & Brook, J.S. (1999). The influence of child and adolescent psychiatric disorders on young adult personality disorder. *American Journal of Psychiatry*, 156, 1529 - 1535.
- Leichtman, M. (2004). Projective tests: the nature of the task. In M. J. Hilsenroth & D. L. Segal (Eds.) (2004). *Comprehensive handbook of psychological assessment* (Vol. 2). Personality assessment. New York: Wiley.
- Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of us. *Educational Measurement: Issues and Practice*, 16(2), 14-16.
- Mellsop, G., Varghese, F. T. N., Joshua, S., & Hicks, A. (1982). The reliability of Axis II of DSM-III. *American Journal of Psychiatry*, 139, 1360-1361.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd.ed.). New York: McGraw-Hill.
- Popham, W. J. (1997). Consequential validity: Right concern-wrong. *Educational Measurement: Issues and Practice*, 16(2), 9-13.
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children – Second Edition* (BASC-2). Bloomington, MN: Pearson.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-8, 13.
- Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice*, 26, 54-60.
- Weiner, I. B. (2004) Rorschach assessment: Current status. In M. Hersen (Editor-in-Chief), M. Hilsenroth & D. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2). Personality assessment (pp. 343-355). Hoben, NJ: John Wiley & Sons.
- Westen, D., Dutra, L., & Shedler, J. (2005). Assessing adolescent personality pathology: Quantifying clinical judgment. *British Journal of Psychiatry*, 186, 227-238.
- Widiger, T. & Samuel, D. (2005). Diagnostic categories or dimensions? A question for the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition. *J. Abnormal Psychology*, 114(4), 494 – 504.
- Wiggins, J. S. (2003). *Paradigms of personality assessment*. New York: Guilford Press
- Zimmerman, M., & Mattia, J. I. (1999). Differences between clinical and research practices in diagnosing borderline personality disorder. *American Journal of Psychiatry*, 156, 1570-1574.