

CE 02 9

Ahora seguimos en Revista de Ciencias de la Educación. Estamos hablando de la problemática de los TES desde la perspectiva del diagnóstico pedagógico y la orientación educativa con los profesores Ramón Pérez Juste y Piedad Granados. Hemos hablado de las características de los TES como instrumento de medida, también de sus problemas sin limitaciones.

Para empezar, queremos destacar que está lejos de nosotros cualquier consideración que pueda llevar a los oyentes a la convicción de que proponemos eliminar la participación del profesor a través de una observación sistemática en el conocimiento de los escolares, objetivo fundamental del diagnóstico pedagógico. Y lejos, asimismo, de considerar que la información recogida mediante TES está libre de limitaciones. En este sentido, es muy comprobable que la revisión más crítica sobre su valía proceda justamente de los expertos, no tanto para quemarlos o prohibirlos, como veíamos ocurrió tiempos atrás, cuanto para utilizarlos con los debidos requisitos de prudencia en las interpretaciones y decisiones derivadas y de formación sólida de los profesionales.

Además, claro está, de mejorarlos mediante los resultados derivados de las investigaciones. Concluíamos la primera parte resaltando que existían limitaciones en estos instrumentos de medida, los TES, centradas sobre todo en el ámbito de la validez. Pero, ¿qué es la validez y en qué consisten estas limitaciones, profesora Granados? Clásica es la definición de que un TES es válido si mide lo que dice medir, y no otra cosa.

Sin embargo, probablemente sea más correcto y completo referir la validez de un TES a su capacidad para alcanzar las metas, los objetivos para los que fue construido. Como estas metas, estos objetivos, son variados, no cabe una única definición de validez. A partir de la edición de las recomendaciones técnicas para los TES psicológicos y las técnicas diagnósticas, y de las normas para las pruebas y manuales educacionales y psicológicos, se suelen presentar como cuatro grandes objetivos o metas.

La primera es la de estimar el resultado que alcanzaría un sujeto en la medida de un universo de situaciones a partir del logrado en una muestra del mismo. Por ejemplo, estimar cuánto tiempo o cuánta ortografía sabe a partir de los resultados de un TES de ortografía. Un segundo caso, una segunda meta u objetivo, sería determinar la posición o lugar ocupado por un sujeto en una variable exterior al TES, bien sea en el momento presente, bien en el futuro o incluso en el pasado.

Estaríamos, en este caso, ante la validez denominada concurrente y la predictiva. Tal sería el caso, por ejemplo, de que a partir de los resultados de un TES de ortografía pudiéramos clasificar al sujeto en una categoría con arreglo a los resultados. O bien, que a partir de los resultados de un TES aplicado al principio de la universidad, se pudieran predecir las probabilidades de éxito académico al final de la carrera.

Por último, inferir el grado en que se posee un rasgo o atributo, una característica o una cualidad, tal como la inteligencia, la emotividad o la introversión. Esto es lo que, técnicamente, se conoce como validez de constructo. Hablaremos, entonces, de cuatro tipos de validez.

Validez de contenido en el primer caso, validez concurrente y predictiva en el segundo, validez de construcción en el último. Bueno, pero, además, los resultados de cada una de ellas no vale para todos los casos y situaciones. Ni para siempre, claro.

Así, la validez de tipo concurrente o la predictiva se refieren a los criterios utilizados y tal como fueron utilizados. Un buen coeficiente de validez predictiva a un año puede ser tan bueno en el caso de medio año o de dos, pero es más fácil que sea diferente en tales circunstancias. Naturalmente, si cambia el criterio o las características de los sujetos, el coeficiente de validez debe calcularse de nuevo.

Esto ocurriría, por ejemplo, si se introducen cambios en el plan de estudio o si se calculó en una población urbana o se desea utilizar en ámbito rural. Si me lo permitís, yo deseo hacer ahora algunas consideraciones que evidencien las dificultades y limitaciones a las que nos estamos refiriendo. Probablemente, el caso menos problemático de todos, al menos en apariencia, sea el de la validez de contenido, de aplicación fundamental a las pruebas de rendimiento escolar.

Precisamente por esa pretendida facilidad, quisiera destacarlo para que se vea que no lo es tanto. En estas pruebas es bastante asequible un muestreo aleatorio o estratificado, ya sea constante o proporcional, para lograr la representatividad en los ítems que integran la prueba. Es decir, si queremos una prueba de rendimiento académico en un área concreta del saber, queremos que haya ítems de los diversos bloques de conocimientos y los diferentes tipos de objetivos.

Bien, hasta ahí la cosa no parece que sea demasiado difícil. Sin embargo, las dificultades surgen en toda su crudeza cuando el redactor de los ítems se pone a la tarea. Y la prueba la tenemos por vía empírica, no por vía de opinión.

Así, por ejemplo, en una investigación realizada en el ámbito de la Universidad Complutense, en que se sometió a análisis la prueba que se realiza para el graduado escolar, se pudo comprobar que la inmensa mayoría de tales ítems se centraban en las categorías más bajas de la taxonomía de Bloom. Es decir, en conocimiento de datos, de hechos y cifras, y en la comprensión en su nivel más bajo, sin apenas relaciones, transferencias o interpretaciones. Por supuesto, muy pocos elementos de aplicación, y menos, por no decir ausencia total, de los del tipo análisis, síntesis o evaluación, que son mucho más complejos.

¿Por qué ocurre esto? Es que los profesores no persiguen o no aspiran a que sus alumnos alcancen estas metas que son importantísimas en la formación intelectual. Bueno, sería lamentable que la respuesta fuera afirmativa. Yo, más bien, creo que se trata de la intrínseca dificultad para formular este tipo de ítems, lo que les lleva a quedarse en los de categorías inferiores.

Ahora bien, si los profesores persiguen todos los objetivos, aún más los complejos y los formativos también, tendríamos que son perseguidos más no evaluados. Y de ahí a que no sean alcanzados, dada la conocida relación entre lo que se evalúa y lo que se consigue, no media más que un paso. Claro, es que ahí está el gran problema.

Encontrar reactivos que evidencien que sea un signo, una manifestación observable de esas características de la persona no observables en forma directa. No sería ese el caso si nos interesara medir su talla o su peso, incluso el número de plaquetas o su presión sanguínea, por ejemplo. Pero esas actitudes o características no son asequibles directamente, sino a través de las manifestaciones que se supone que las evidencian.

Bunker diría a esto que es una hipótesis. Esta limitación que ha resaltado Piedad no es exclusiva de la validez de contenido. Si me apuran, es superior en las variables de la personalidad en las que este tipo de validez es menos utilizado.

Véase si no un caso de ítem típico en un cuestionario de personalidad. Por ejemplo, si usted entra en un local cerrado donde se está dando una conferencia, todos los asientos están ocupados menos la primera fila, ¿usted subiría a esa primera fila o no? En la hipótesis del investigador está que si uno contesta que sí, es que se trata de una persona con gran capacidad o dominio de sí mismo, con gran seguridad. Y si dice que no, se entiende que es porque se trata de una persona tímida.

Bien, pero el mismo autor de este ítem, concretamente Allport, dice él mismo. Bien, puede ocurrir que diga que no sube a la primera fila y sea simplemente por una manifestación de respeto hacia los demás, por no distraerles. Y si se queda a la última fila, bien puede que no sea porque se trate de una persona tímida o insegura, sino que, tal vez, va a irse antes de que acabe la conferencia y no quiere molestar.

Como se ve, una cosa es la suposición que está prevista en la hipótesis del investigador y otra cosa es realmente lo que la persona, en concreto, en esa situación concreta, opina y hace. Bien, pues este hecho en concreto, esta falta de validez aparente, es la que puede alterar los resultados de las pruebas de diagnóstico. Ahora, si de la validez de contenido pasamos a la validez concurrente, el problema al que acabamos de hacer alusión se extiende también a los criterios.

Es decir, se extiende a esas variables exteriores al test que se utilizan como garantes de su valía para determinar la validez. De su calidad y de la calidad de su medida se deriva el nivel y la calidad misma de la validez concurrente. Piénsese en un ejemplo.

Si el criterio de validez para un test de lectoescritura es el juicio de expertos en ese campo, en la detección de anomalías de este área, la calidad del diagnóstico que, una vez validado el test, pueda realizarse con él, dependerá, en gran medida, de la calidad del juicio de tales expertos. Pues bien, en este sentido, se han hecho estudios, tanto con test proyectivos como psicométricos, por ejemplo, los conocidos Rothschild o MMPI, como son los estudios de Rosen,

de Edwards y de Abbott, de los años 58 y 73, respectivamente, y los resultados, lamentablemente, son un tanto desalentadores. Así, por ejemplo, con el MMPI, uno de los más conocidos test psicométricos del área de la personalidad, se identifica un bajísimo tanto por ciento de pacientes.

Concretamente, los resultados hablan de en torno del 50%. Por otra parte, las clasificaciones derivadas de los resultados de los test no concuerdan sustancialmente con las que derivan de la teoría que los sustenta. Asimismo, hay ocasiones en que se entremezclan diferentes grupos de anormalidades cuando los resultados de los test se les aplican métodos empíricos para la formación de tales grupos.

El caso más concreto es una investigación en que se aplicó el análisis de clúster y donde las categorías hechas con este tipo de análisis no concordaban, se entremezclaban los diversos grupos con arreglo a cómo deberían haberse producido. Claro está que las razones que explican estos hechos pueden ser muy variadas. Los fallos pueden ser debidos a los test o a las características de los sujetos, como veíamos con anterioridad.

Pero también puede ocurrir que lo que no tenga consistencia es la teoría. Las clasificaciones teóricas utilizadas, los constructos, o incluso que las técnicas estadísticas aplicadas no se utilicen con el debido rigor y respeto a las exigencias de los modelos. Sí, es muy interesante respecto a la posición que en torno a estos temas está tomando el movimiento conocido como evaluación conductual.

Sus defensores resaltan que la inconsistencia de los sujetos en las respuestas tal vez no sea tal inconsistencia, sino más bien el fruto de una cierta especificidad en función, sobre todo, de la situación en que se produce la conducta, lo que se traduce en una superior complejidad y en la dificultad de alcanzar categorías netas, claras y precisas. Puede que haya un poquito de todo, ¿verdad? Pero, de ser ciertas las objeciones de la evaluación conductual, tal vez sea preciso iniciar una cierta reforma en los planteamientos de los test que den cabida a la especificidad de las situaciones. En ese sentido quisiera señalar cómo en el inventario de adaptación social que constituyó mi tesis doctoral se referían los ítems, básicamente los mismos, a siete situaciones sociales diferentes, buscando justamente esa especificidad.

Pues bien, los resultados que yo logré parecen confirmar los planteamientos anteriores, ya que las correlaciones que se daban entre las distintas situaciones, aunque suficientemente elevadas, entre ellas oscilaban entre 0.444 y 0.724, distan, como está, mucho de ser perfectas. Y, sin embargo, y para que se vea cuán complejo es esto, Tiller, por ejemplo, nos habla de un informe, de una investigación sobre la estabilidad de los rasgos, en la que se apreciaron unas predicciones correctas en un 66% sobre un total de 882. Quisiera resaltar que en este punto de la estabilidad está uno de los elementos base de la evaluación tradicional en la que se apoyan los métodos psicométricos.

Si la estabilidad se pone en tela de juicio, es necesario meternos en el campo de la evaluación conductual. Bueno, han hablado ustedes, profesor Pérez Juste, profesora Granados, de las

limitaciones y problemas respecto de la validez de contenido y de la validez concurrente. Pero antes la profesora Granados nos habló de otros tipos de validez.

¿Cuál es la situación respecto de lo que ustedes han denominado validez predictiva? En este punto, las limitaciones son dobles. Por una parte, se apunta a la misma problemática de los criterios, complicado con el hecho de que, en este caso, en muchas ocasiones, no se puede utilizar los últimos o absolutos debiendo conformarse con los intermedios, lo que no deja de ser una limitación. Por otra parte, la enorme complejidad de la conducta humana, lo que se traduce en la incidencia de muchas variables en el aspecto comportamiento a explicar, esto es, en nuestro caso, a predecir.

Esto parece obligar a utilizar modelos predictivos complejos basados no en una variable predictiva o test, sino en un conjunto de estas para que la predicción del criterio sea eficaz. Sí, los estudios de Rosen y de Edwards, a los que ya hemos aludido, evidenciaban ya esas limitaciones. Hay que resaltar que las limitaciones se refieren no a conductas de poca monta, sino incluso a las más extremas o drásticas, como puede ser incluso hasta el mismo suicidio.

Sin embargo, es preciso reconocer una cierta eficacia en cuanto a la predicción para grupos. Es decir, menos eficacia en la predicción individual, más eficacia para predecir para grupos. Claro que a esto habría que añadir un nuevo elemento de dificultad adicional.

Piénsese, por ejemplo, que todos los resultados de validez, tanto concurrente como predictiva, son resultados empíricos. Es decir, resultados que se refieren a las respuestas dadas sobre una hoja de papel entre las puntuaciones del test y el criterio. Pero, claro, uno puede pensar, se puede suponer, sin más, que en una situación real, el sujeto obraría, como afirma, cuando se le pregunta sobre un papel.

Con todo, y a pesar de esto que acabamos de decir, las limitaciones no son exageradas cuando los test se utilizan sólo para diagnosticar validez concurrente o sólo para predecir. Donde el problema se manifiesta en términos más agudos, y la explicación que la venimos dando a lo largo de este diálogo, es cuando, a partir de resultados, se pretende tomar ciertas decisiones pertinentes, como, por ejemplo, clasificar, diseñar estrategias o prescribir intervenciones. Pero claro está que estas decisiones son lo que de verdad interesan a cualquier pedagogo.

Saber, simplemente, el resultado de la inteligencia de un chico o las dificultades ortográficas, no resuelve nada si no sirve después para tomar esas decisiones. Bueno, nos queda hablar del cuarto objetivo, de la validez de constructo. ¿Cuál es la situación de cara a este cuarto objetivo de validez? Profesor Pérez Juste.

Bueno, aquí es necesario resaltar que en la validez de constructo no llegamos nunca a un valor numérico, como los demás tipos de validez concurrente o predictiva, sino, más bien, a una progresiva acumulación de evidencias científicas que parecen confirmar la teoría sobre la que se edifica el constructo, o que, por el contrario, la ponen en tela de juicio. Pues bien, hay un hecho curioso que evidencia las limitaciones a las que estamos refiriéndonos. Cuando un

mismo rasgo o característica se mide con diferentes tes, las correlaciones alcanzadas con sus resultados son muy bajas.

Tiller nos informa de valores en torno a 0,36 para una investigación realizada por Vernon sobre intervención. Piénsese que lo que se está midiendo es una misma cosa con distintos tes. El resultado parece desalentador.

Sin embargo, y esto todavía más chocante, hay veces en que diferentes rasgos, por ejemplo, pongamos inteligencia y emotividad, medidos con el mismo método o con la misma técnica, dan lugar a correlaciones de ese estilo e incluso superiores. Estos hechos levantan la sospecha de si no será justamente el método de medida lo que produce las correlaciones, y no el rasgo que se supone su yace. Y piénsese qué conclusiones se podrían obtener si midiendo inteligencia y timidez llegáramos a una validez de 0,50.

¿Cómo se podría llegar siquiera a interpretar este hecho? Podríamos concluir nuestra intervención señalando que, dada la complejidad de lo que se desea medir, dados los múltiples enfoques utilizados, dadas las limitaciones metodológicas, no debe considerarse extraño que existan estas contradicciones y estas limitaciones. Que tanto los datos procedentes de la observación del profesor, cuanto los originados en la aplicación de test, deben tenerse en consideración a la hora de diagnosticar, analizando congruencias e incongruencias, y que conviene recoger información con técnicas e instrumentos diversos, que es necesario que el diagnóstico sea acumulativo y siempre provisional, huyendo de todo encasillamiento y de toda rigidez. Y que hay que ser enormemente prudente en las interpretaciones y decisiones, y profundamente respetuosos con la persona humana sobre la que se centra el diagnóstico.

Los profesores Ramón Pérez Juste y Piedad Granados nos han hablado de las características de los test como instrumentos de medida que permiten alcanzar resultados superiores a los alcanzados a través de la observación, pero también de los problemas y limitaciones respecto de la validez de los test. Quede claro que, como decía la profesora Granados, ni el diagnóstico pedagógico se identifica con los test, ni estos son los únicos instrumentos utilizados, ni se pretende anular o limitar la participación del profesor. El diagnóstico debe ser acumulativo y siempre provisional, siendo prudentes en las interpretaciones y decisiones, respetando siempre a la persona sobre la que se centra el diagnóstico.

Y ahora, otras cosas. Los próximos días, 16 y 17 de noviembre, en sesiones de mañana y tarde, el día 16, y de mañana, el día 17, se va a celebrar, bajo la patrocinio de la Real Academia de Ciencias Morales y Políticas, un seminario sobre Pedagogía Universitaria. Este seminario está organizado por el catedrático de la Universidad Complutense, Víctor García-Oz, que es catedrático de Pedagogía Experimental y Diferencial.

La Universidad Complutense, aprovechando que en él va a participar como ponente el profesor Ramón Pérez Juste, director del Departamento de Pedagogía Experimental y Orientación de la UNED, departamento al que hoy dedicamos la Revista de Ciencias de la Educación, le preguntamos por los objetivos perseguidos en esta reunión científica. ¿Cuáles son estos

objetivos, profesor Pérez Juste? Pues, mire, básicamente, el tema es el que, desde hace muchísimos años, no hay ninguna obra de conjunto de tipo Pedagogía Universitaria, ha habido algunas obras parciales sobre algunos aspectos, y el profesor García-Oz, muy preocupado desde ya hace años por este tema, se ha propuesto que nos reunamos unos cuantos profesores como ponentes y otros como participantes para reflexionar globalmente sobre el tema de la Pedagogía Universitaria. Y del fruto de esa reflexión, muy probablemente saldrá un tratado de Pedagogía Universitaria que se puede editar y, digamos, que esté al día en torno a los temas que nos preocupan.

¿Y cuáles son estos profesores que van a intervenir en este Seminario de Pedagogía Universitaria? Pues, entre los ponentes, se encuentran profesores de diversas universidades españolas, por ejemplo, abre fuego, digamos, el profesor Ibañez Martín, de la Universidad Complutense, con un tema relacionado con sus preocupaciones, las dimensiones formativas del quehacer universitario. Tenemos también entre nosotros ese día al profesor Medina Rubio, catedrático de Oviedo, al profesor Vázquez Gómez, que es catedrático de la Universidad Complutense, al igual que el doctor Arturo de la Orden. Participarán también dos catedráticos, uno de Barcelona, el profesor Quintana Cabanas, y el profesor Rodríguez Lligue, de Salamanca, dirigirá, como digo, el profesor García Hoz, y también tendremos entre nosotros a una profesora titular de Pedagogía Experimental de la Complutense, la profesora Díez Ayue.

Naturalmente, también intervendré yo en una de las ponencias. ¿Y cuál es esta ponencia, profesor Pérez Juste? Mi ponencia lleva el título genérico de Programación de Actividades Universitarias, y habría que decir que se inscribe en el marco de las tres ponencias anteriores, que comienza con lo que he señalado del profesor Ibañez Martín, dimensiones formativas del quehacer universitario, se acerca un poquito más a la realidad con los fines formativos de la universidad que presentará el profesor Medina Rubio, y me precede inmediatamente el doctor Vázquez, que nos presenta Masificación y Calidad Universitaria, la suerte de la universidad entre el igualitarismo y el desarrollo. Después de mí vendrán otros profesores que ya se preocupan de cuestiones más cercanas, incluso como puede ser la evaluación.

Al final, terminaremos con una mesa redonda donde el objetivo fundamental será llegar al índice de ese posible futuro Tratado de Pedagogía Universitaria. ¿Y antes de este tratado seguirá alguna publicación a estas sesiones del seminario? Es muy posible que, antes de que pueda llegar como tal el tratado, salga una publicación con las ponencias que los diversos ponentes presentamos. Inicialmente, cada ponente solo va a tener media hora para exponer su tema, después habrá un coloquio, pero no hay ningún límite en cuanto a la extensión del documento escrito que presente.

Si los documentos tienen la entidad suficiente, desde el punto de vista del coordinador, es posible que la Real Academia o alguna otra entidad lo patrocine y en adelante pudiera salir ese otro magno tratado, pero inicialmente sí puede que haya esa publicación. Bueno, y ya por último, ¿a este tipo de seminarios puede asistir cualquiera o tiene que ser un experto en temas de pedagogía? Inicialmente, los participantes tienen dos categorías distintas. Uno son el de los

ponentes, que estamos invitados y que se ha solicitado a nosotros que estemos presentes en todas las ponencias, con el fin de que luego pueda darse un diálogo entre todos ellos.

Pero también, en principio, queda abierto para cualquier profesor, no solo de pedagogía, y también incluso para los alumnos. Lástima que los alumnos de la UNED, que están un poco lejos, no puedan venir, pero si algún alumno de Madrid, y sobre todo de los centros más cercanos, quisiera acercarse, pues que sepa que las sesiones van a comenzar el día 16 a las 10 de la mañana, en la Real Academia de Ciencias Morales y Políticas, y que terminan por la tarde alrededor de las 8. Y al día siguiente, el sábado 17, comenzamos un poquito antes, comenzamos a las 8 de la mañana, con intención de poder acabar al mediodía. Esto ha sido todo en Revista de Ciencias de la Educación.

En el capítulo de Orientaciones, las profesoras Elvira Repeto y Araceli Sebastián, se han dirigido a los alumnos de Teoría y Procesos de la Orientación, Especialidad de Orientación Escolar, y a los alumnos de Pedagogía Experimental I, respectivamente. El director del Departamento de Pedagogía Experimental y Orientación, Ramón Pérez Juste, y la profesora de Diagnóstico Pedagógico y Técnicas de Orientación, Piedad Granados, nos han hablado de la problemática de los test. El próximo jueves, elaborarán la Revista de Ciencias de la Educación los profesores del Departamento de Metodología Educativa.

Serán temas relativos a técnicas de diagnóstico y evaluación educativa, la didáctica actual y tecnología educativa.