

## Epílogo

### *Reflexiones en torno a la validación empírica del modelo de medición de la teoría clásica de tests (TCT)*

“... En la práctica, sin embargo, el constructor del test no puede estar seguro de crear mediciones perfectamente paralelas, y tampoco pueden obtenerse las puntuaciones verdaderas de los examinados. Entonces, ¿cómo puede determinarse el coeficiente de fiabilidad para un conjunto de mediciones ? La respuesta es que el coeficiente de fiabilidad para un conjunto de puntuaciones nunca puede ser determinado exactamente; puede, sin embargo, ser *estimado* para una muestra dada de individuos respondiendo a una muestra dada de ítems de tests.”

(Crocker y Algina, 1986; 131)

Para efectuar una medición psicológica necesitamos simultáneamente de una teoría psicológica sobre la variable a medir y una teoría psicométrica sobre como puede ser medida. Tanto la teoría psicológica como la teoría psicométrica, en la medida en que alcanzan cierto grado de elaboración y sistematización, pueden expresarse por medio de un modelo formal. Un modelo formal puede verse como una teoría o un fragmento de teoría, expresado en algún lenguaje formal, generalmente matemático, que establece un conjunto bien definido de relaciones sintácticas (las relaciones formales de unos constructos con otros) y semánticas (la relación entre los constructos y la realidad) y permite efectuar deducciones precisas que puedan someterse a contraste empírico. Un modelo formal es así un conjunto estructurado de proposiciones formales justificado por su capacidad para mejorar nuestra comprensión y capacidad de pronóstico de la estructura y funcionamiento de alguna parcela acotada del mundo real.

Los modelos formales de medición en psicología tienen el referente común de un sustrato de fundamentación matemática, especialmente estadística, que contribuye a definir sus supuestos iniciales y sus deducciones. Esa teoría matemática, particularmente referida a la teoría de la probabilidad, se ha tomado tradicionalmente por universalmente cierta y no se considera objeto de discusión

empírica, es decir, no se toma en cuenta la posibilidad de que ningún conjunto de datos reales pueda rebatirla, de modo que, cualquier eventual discrepancia entre los datos y esta teoría se desprecia como irrelevante, rechazando los datos y no la teoría. Por supuesto este modo de proceder puede desperdiciar ocasionalmente la oportunidad de efectuar descubrimientos interesantes, pero esta posibilidad ni siquiera se insinúa, de modo que este no es el objeto de discusión. Solo muy recientemente la universalidad de los modelos estadísticos para toda clase de fenómenos psicológicos ha comenzado a ponerse en duda.

Más allá de los supuestos probabilísticos y estadísticos no expuestos a contraste, los modelos hacen *supuestos específicos*, especialmente acerca de las distribuciones de las puntuaciones de las mediciones psicológicas, del acto de medición, del comportamiento del objeto medido, del significado de la respuesta y de su interpretabilidad. Establecido un conjunto de supuestos mínimo y no contradictorio se obtienen deducciones formales que habrán de tener en algún punto consecuencias observables relevantes. La naturaleza de este modo de razonar es la siguiente: Si los supuestos son ciertos, entonces, las conclusiones observables obtenidas a partir de deducciones formales mediante las reglas estadísticas y matemáticas antes aludidas, se verificarán (o al menos, razonablemente, no serán falsadas) por los datos. La calidad del modelo habrá de ser juzgada por su capacidad de ofrecer información cierta y útil sobre el mundo real, sobre las

mediciones psicológicas reales efectuadas con sujetos reales en contextos reales. Debido a que realmente ese conjunto mínimo de supuestos contiene toda la información suficiente que genera las deducciones, frecuentemente se emplea la palabra 'modelo' en este contexto para referirse precisamente a ese núcleo mínimo de proposiciones.

Lord y Novick (1968) razonan claramente sobre esta importante cuestión en los siguientes términos: "Si esas asunciones adicionales [se refieren a los supuestos específicos que acabamos de mencionar] son válidas, entonces las derivaciones lógicas (matemáticas) a partir del modelo describen correctamente propiedades relevantes más complicadas de datos de puntuaciones de los tests. Esto es precisamente lo que queremos. Si esas asunciones adicionales no reflejan precisamente el verdadero estado de cosas, entonces las inferencias desde esos modelos *pueden* ser incorrectas y posiblemente realmente engañosas. Por esta razón nuestros modelos deben ser establecidos y empleados con mucha prudencia. Debemos resistir la tentación de hacer supuestos simplemente porque deseamos obtener ciertos resultados y fórmulas.

La determinación de la validez o legitimidad de un modelo matemático debe ser basada normalmente en dos tipos de estudio separados. El primer tipo debería demostrar la validación empírica del modelo. Esto se cumple haciendo predicciones desde el modelo y viendo si esas predicciones son sostenidas por los datos. El segundo tipo de estudio

debería aplicar el análisis lógico y matemático a los supuestos del modelo de tal forma que su validez o falta de validez se haga más obvia y visible.” (Lord y Novick, 1968)

Primero, los supuestos no deben ser irreales. Es posible escribir diversos conjuntos mínimos de supuestos irreales, que no se corresponden con las mediciones que hacen o pueden hacer los psicólogos en el mundo real, y deducir elegantemente sistemas formales tan admirables como alejados del mundo real. Esto es interesante desde un punto de vista formal, pero es de menos ayuda desde un punto de vista aplicado. El problema de los modelos de medición no está del lado de las matemáticas. La cuestión reside en que los modelos, para que ayuden a la psicología a medir sus variables de interés, han de tener capacidad de explicar la realidad y permitir pronósticos contrastables sobre cuestiones no irrelevantes. Y han de ser aplicables, en un sentido práctico. Formalmente, los supuestos pues han de ser “realistas”, pero sobre todo han de llevar a deducciones contrastables relevantes.

Segundo, por tanto, la calidad de los modelos formales dependerá de su validación empírica, de su capacidad para reflejar lo que sucede con las mediciones reales que efectúan o pueden efectuar realmente los psicólogos.

A mi juicio, hay algo inquietante en las palabras de Lord y Novick que trasluce una sospecha crítica sobre un estado de cosas que ha resultado frecuente en la teoría de tests. Si un modelo ha de ser juzgado por su validez empírica, es

decir, por el grado en que datos reales sostienen sus predicciones, entonces ¿por qué debemos resistir la tentación de hacer supuestos simplemente para obtener determinados resultados y fórmulas?, y, ¿por qué nuestros modelos han de ser empleados con mucha prudencia?

El proceso para establecer un modelo de medición debería suponer tres pasos. Primero, establecer los *supuestos mínimos generadores* del modelo, que, en segundo lugar, darían lugar a las *deducciones* pertinentes y, que, en tercer lugar, tales deducciones deberían ser expuestas a *contraste empírico*. Aunque sin duda el proceso de trabajo no sería tan lineal, los tres pasos esenciales habrían de ser estos.

Como consecuencia de estos tres pasos el modelo sería o bien rechazado como incorrecto, si las deducciones del modelo no fueran soportadas por los datos, o bien aceptado como provisionalmente correcto si hasta el momento los datos no lo han rechazado o no existe otro modelo alternativo con mejor ajuste a los datos.

El mecanismo básico de la lógica de decisión científica actúa como un “*modus tollens*”: Si no son ciertas las consecuencias observables no son ciertos los supuestos. Obsérvese que no puede decirse que si sí son ciertas las consecuencias los supuestos son ciertos, únicamente que son compatibles con esas consecuencias.

Bien, si este es el proceso ¿por qué no hacer los supuestos que se desee? ¿acaso no van a ser contrastados a través

de deducciones expuestas a los datos? Si un modelo es soportado por los datos ¿por qué usarlo con prudencia? y si no es soportado por los datos ¿por qué usarlo con prudencia?.

La advertencia de Lord y Novick trasluce un estado de cosas característico en TCT, y parcialmente en otros modelos. La discusión de los modelos debería tener tres componentes: los supuestos, las deducciones y los datos de contraste que los apoyen. En TCT la situación es sorprendente. Los modelos tienen supuestos. Los modelos tienen deducciones y más deducciones. Pero los modelos no tienen contraste empírico.

Lord y Novick también han sido explícitos en algún punto acerca de esta naturaleza anómala de la TCT. En sus propios términos "... el modelo de teoría clásica de tests es una tautología más que un modelo o una teoría. Con esto queremos decir que el modelo debería mantenerse con respecto a cualquier conjunto dado de datos. Así, para la mayoría de propósitos, la validez empírica y utilidad del modelo residen en el supuesto de independencia experimental lineal y la disponibilidad de mediciones paralelas. El supuesto de independencia experimental lineal es ciertamente un supuesto relativamente débil. En muchas aplicaciones de tests este es un supuesto que generalmente estamos preparados para hacer al menos para series cortas de observaciones. Para propósitos de estimación, usualmente necesitamos un mínimo de sólo dos

mediciones paralelas, y en el campo de la medición con tests mentales deseamos suponer que podemos obtener al menos dos mediciones aproximadamente paralelas..." (Lord y Novick, 1968; 48). Una tautología soportada por cualquier conjunto de datos no puede ser rechazada empíricamente por su propia naturaleza, y, por esa razón, no se la puede considerar un modelo en el sentido científico habitual.

Hace más o menos un siglo desde que nació el modelo lineal aditivo que fundamenta la TCT. Se han construido literalmente decenas de miles de instrumentos de medida bajo esta teoría. Se han calculado sus fórmulas miles y miles de veces en todo el mundo durante décadas. Y sin embargo no se dispone de una valoración de la misma en términos de validez empírica. Esa carencia de validación empírica fundamenta el temor a los supuestos y prudencia en el uso de modelos que traslucen Lord y Novick.

La validación empírica de un modelo es esencial. Obsérvese que, teóricamente, es posible construir virtualmente infinitos modelos técnicamente perfectos cuyo único defecto sea su incapacidad para describir y predecir el mundo real. Un buen conjunto de supuestos no contradictorios entre sí puede generar con un poco de cálculo, álgebra y paciencia un conjunto de deducciones. Si alguna vez hubiera instrumentos de medida como los que supone el modelo, y sujetos como los que supone el modelo y condiciones como las que supone el modelo entonces el modelo sería empíricamente válido. Mientras el

programa de investigación espacial encuentra la Galaxia Psicometría donde quizás habitan esos seres, esas condiciones y esos instrumentos tan solo tenemos un juguete matemático, no importa cuan elegantes sean sus supuestos y deducciones, cuan rigurosa sea su notación o cuan general sea su planteamiento.

Las buenas presentaciones formales de los modelos -por ejemplo la que hacen Lord y Novick (1968) de la Teoría Clásica de Tests- son extraordinariamente abstractas y optan deliberadamente por un nivel de definición y de exposición sintáctico, es decir, de elementos y relaciones formales, frecuentemente con alto grado de generalidad, con economía de supuestos y elegancia en la deducción. Pero la cuestión clave es la conexión semántica entre modelo y realidad y el contraste empírico del modelo a través de sus consecuencias mediante esa conexión semántica.

En mi opinión la psicometría está demasiado llena de sofisticados juguetes matemáticos que se miran al ombligo. Describen mundos hipotéticos más o menos alejados de las mediciones reales de psicólogos reales con sujetos reales. En la misma medida su utilidad para el trabajo real se vuelve dudosa. Muchos de estos juguetes matemáticos no resisten la pregunta más elemental que se debe formular a una teoría: ¿dónde están sus pruebas? ¿dónde están los datos de contraste que apoyan el modelo?

Aquí he repasado algunos puntos básicos de la TCT bajo una perspectiva crítica. La TCT es esencialmente una teoría de la “fiabilidad” adecuada para tests de aptitudes colectivos de lápiz y papel que, conviene recordar, son muy corrientes pero no los únicos. Puede verse como uno de aquellos grandes edificios de cristal y metal que se construían a principios del siglo XX. Un ejercicio de la ingeniería de la medición psicológica que se sostiene graciosamente sobre unos pocos nervios, supuestos y métodos, bien presentados aunque, en mi opinión, débiles.

Spearman fue el padre de la TCT al principio del siglo XX. En la primera mitad del siglo la Teoría Clásica de Tests creció, se fortaleció, y se sistematizó en la obra de Gulliksen (1950), una presentación lucida a la que todos los trabajos posteriores aun debemos comprensión. Después, con la obra de Lord y Novick (1968) vendría la puesta de largo, con una formalización matemática más precisa y elegante de poca repercusión semántica (en lo que a Teoría Clásica de Tests se refiere) pero considerable calidad sintáctica. Las ideas, las fórmulas y las deducciones se repiten de manual en manual. Hay una tendencia a no entrar en su discusión. Ni siquiera los retazos críticos que aparecen en la “línea oficial”, como la cita de Lord y Novick que antecede a estas páginas, suelen ser mencionados. Escribir un manual de teoría clásica de tests siguiendo la línea clásica es un ejercicio de didáctica. Desde luego este es un camino “seguro” para los autores. Parece que el punto de vista medieval de ampararse en las autoridades

sigue siendo académicamente eficaz: “Lord y Novick o Gulliksen dixit” ¿Para qué repensar el problema?. En la medida en que se hacen innovaciones, o se discute críticamente el camino comúnmente admitido, se asumen riesgos. Esencialmente el riesgo de equivocarse que es un riesgo inherente a la creación intelectual y científica. Mi objetivo ha sido pensar esta vieja teoría que no se sabe si es una teoría. *Pensar y no solo aprender*. Además de intentar exponer los conceptos importantes.

Creo que el modelo clásico de tests es hoy una buena base para empezar a discutir y razonar cuestiones principales de la medición en Psicología. Me parece que, si se mira con ojos curiosos, es un buen modelo por las preguntas que deja planteadas. Por cierto que muchas de esas preguntas fundamentales no son contestadas sino soslayadas por la teoría de la respuesta al ítem, que es la teoría dominante. Sin embargo, no me parece que la TCT sea un modelo que pueda presentarse hoy en día como materia de estudio sin una reflexión crítica. A mi me resulta impresionante que puedan seguir escribiéndose manuales sobre Teoría Clásica de Tests como si fuera una teoría correcta o adecuada o aceptable, que el lector no tiene más que aplicar (aunque se deje para el estudiante avanzado profundizar en las matemáticas vía apéndices o problemas de demostración). Se argumentará que razones didácticas llevan a no entrar en complicaciones. La verdad es que no comprendo bien que ninguna razón que lleve a no reflexionar por si mismo pueda calificarse de didáctica. De

lo dicho se desprende que el lector deseoso de ejercicios de credulidad puede encontrar mejores sitios donde estudiar la TCT.

Aquí no he perseguido la elegancia matemática y el nivel de discusión sintáctico que caracterizan a la obra de Lord y Novick (1968). Allí están todas las piezas principales de la TCT orquestadas con el rigor y la calidad que ofrece un entendimiento sistemático. Por el contrario, en contra de lo que ya es tradicional en los trabajos aparecidos en los últimos años, he vuelto intencionadamente a un nivel de presentación más sencillo, como el que utilizó Gulliksen (1950) y popularizó entre nosotros el profesor Yela (1984). Después de todo no he encontrado que ese nivel de formalización tenga ninguna repercusión práctica sobre el modo en que se puede utilizar la teoría.

Mi relación personal con la TCT fue una pasión difícil desde el primer momento. La primera vez que asistí a una clase de psicometría el profesor se empeñó -sin previo aviso- en suponer que el error de medida se *sumaba* a la puntuación verdadera y empezar a deducir fórmulas y más fórmulas como si tal cosa. Pregunte por qué se *sumaba* y no obtuve una respuesta satisfactoria. Después el profesor comenzó a *sumar* ítems acertados, lo que supuestamente daría una medida de la capacidad inteligente de un sujeto. Pregunte cómo una capacidad compleja, lógica, estructurada, podía reflejarse bien con un procedimiento tan burdo como sumar

aciertos a problemas más o menos heterogéneos que ponían en funcionamiento diversas capacidades, conocimientos, connotaciones etc. a diversos niveles. Me parecía (y me sigue pareciendo) una parodia de un intento de comprensión y explicación científico del pensamiento de la comprensión, del razonamiento, de la inteligencia. Y, desde luego lo más alejado de una medición científica precisa. No recuerdo que me contestaron, pero si recuerdo que no volví a aquella clase en el resto del curso (excepto los días de examen por razones plebeyas bien conocidas).

Las cosas principales de la TCT las fui aprendiendo más tarde, poco a poco, estudiando los apuntes del profesor Mariano Yela en los que hemos aprendido y entendido la psicometría la mayoría de los psicólogos españoles de mi generación (muchas gracias, Profesor). Después, he intentado aprender y todavía sigo intentado aprender de los grandes manuales clásicos mencionados tantas veces en el texto, y de otros más recientes, y también, por supuesto, de algunos manuales excelentes producidos aquí, como los de la profesora Santisteban entre otros. Quizás mis colegas sepan y puedan justificar mejor que yo los supuestos y deducciones que hacia mi primer profesor de psicometría, ante los que, por cierto, sigo siendo esencialmente escéptico después de haber leído muchas respuestas aceptables. Afortunadamente mis estudiantes más inquietos continuarán ayudándome a comprender un poco más las preguntas de fondo que todavía esperan respuesta. Todavía creo que en la medición psicológica -y en buena

parte de la psicología- las preguntas todavía valen más que las respuestas disponibles.

Una lectura cuidadosa de los interrogantes críticos que he ido formulando a lo largo de las páginas anteriores ha de llevar a un estudio y uso reflexivo y prudente de la psicometría y de los tests en general. La medición psicológica es un instrumento fantástico que lleva a poder operar de un modo más sencillo sobre una realidad compleja y que puede ayudarnos de un modo esencial a mejorar nuestra comprensión de las personas. Si creo que es necesaria una lectura crítica de la TCT, y en general de la psicometría, es precisamente porque creo profundamente en el interés psicológico de la medición psicológica.

Tengo la sensación de que la medición en Psicología es todavía una actividad incipiente, a pesar de muchas páginas llenas de un siglo de matemáticas. Mi punto de vista es particularmente crítico. Pero hay que envolverse en el manto de un academicismo estéril para estar satisfecho del actual nivel de desarrollo de la medición psicológica, incluida por cierto la Teoría de la Respuesta al Ítem, de la que no hemos tratado aquí. Hay desarrollos rigurosísimos edificados sobre supuestos irreales. Mucho análisis sofisticado de datos muy pobres. Mucho decimal donde no se sabe bien que es lo principal. Hay un inmenso edificio métrico deshabitado de psicología. Y probablemente el problema esencial esta de lado de la psicología más que del método.

En conjunto la TCT sirve tanto para evaluar la calidad métrica de un instrumento como para enmascarar la ignorancia psicológica acerca de qué significan las puntuaciones de los tests (cuando significan algo). Mi opinión es que la teoría de la fiabilidad clásica no sirve y, en general, como tal teoría de la fiabilidad, debería ser abandonada. No veo evidencia de que la fiabilidad sirva para algo excepto para reducir la validez -al menos en algunas de sus formas-. Y la validez si que es importante. Es lo esencial. En mi opinión una teoría restrictiva de la validez, como la que he esbozado, podría ser útil y contribuir a despejar el confuso panorama de los tests. Sin embargo, en términos latos, tal como ahora se practica la teoría de la validez, es más bien un ejercicio académico de mucho rigor formal y poco rigor real.

Más allá de esto, creo que los procedimientos actuales que producen supuestas mediciones psicológicas deben ser seriamente revisados. La mayor parte de la medición psicológica descansa sobre un proceso elemental y a la vez desconocido: ¿cómo contestan las personas a las preguntas? ¿qué determina su respuesta? ¿por qué los sujetos contestan firmemente preguntas que por su naturaleza son ambiguas? ¿qué métodos utilizan los sujetos para engañar a los psicólogos y que efectos producen sobre las respuestas? ¿qué saben los sujetos de si mismos y en qué medida influye en lo que contestan? ¿Qué procesos iguales y qué procesos distintos están implicados en cada ítem? Para muchas de esas cuestiones

se abrieron líneas de investigación y se puede esperar una respuesta empírica. Tomadas conjuntamente esa clase de interrogantes hace que sienta perplejidad antes de creer que un mismo proceso tan elemental como preguntarle al sujeto (y por, lo general, que ponga algunas cruces aquí y allá) permita “medir” igualmente cientos de constructos inobservables, supuestamente fiables y validados, absolutamente heterogéneos y que, de algún modo habrá que admitir que se refieren todos al mismo único sujeto. Algo no encaja.

La psicometría, y los tests, en general, han dotado a la profesión de métodos prácticos, de modos de hacer profesionales y han contribuido a su estatus social, -y, a veces, desafortunadamente a su descrédito, más por malos usos que por defectos de las teorías métricas de fondo-. Paradójicamente en algunos contextos los tests son de lo mejor, más probado y bien conocido que utilizan algunos psicólogos. En un sentido positivo debe hacerse un esfuerzo por encontrar nuevos marcos psicológicos y psicométricos para el desarrollo e interpretación de instrumentos de medición psicológica. La medición psicológica es necesaria e irrenunciable, en términos prácticos se trata de hacerlo tan bien como nos sea posible, pero sin renunciar a mirar un poco más lejos.

Algunos comienzan los manuales de psicometría con la famosa y brillante cita de E. L. Thorndike según la cual



“todo lo que existe, existe en alguna cantidad” (por ejemplo, Crocker y Algina, 1986, pag. 3). A mi me parece oportuno acabar estos apuntes con algunas preguntas inquietantes: ¿Qué es lo que existe y lo que no existe de los constructos latentes que, se supone, medimos? ¿Qué es lo que existe y lo que no existe de las puntuaciones verdaderas? Porque si todo esto “no existiera” ¿cómo va a existir en cantidad ninguna?

Imaginemos por un momento un nuevo constructo inobservable, que pudiéramos saber cierto que no existe. ¿Qué problema habría en escribir un puñado de items medianamente consistentes y medianamente relacionados con otros supuestos constructos inobservables tal que dispusiéramos de una aceptable medida de nada?

El lector puede imaginar ahora el constructo latente psicológico más absurdo. No importa cuan disparatado sea. No importa si no sabe si es un proceso o un estado. Ni tampoco si es un proceso o un estado de qué. No importa que no sepa si es conducta o si no lo es, o si está relacionado de algún modo con la actividad eléctrica o bioquímica del cerebro o no lo está. No importa que no sepa donde empieza y donde acaba, y tampoco importa si no sabe muy bien como se relacionará con otros inobservables más o menos conocidos. (Después de todo, todo esto tampoco se sabe de la mayoría de las cosas que medimos). Pues bien, no hay más prueba de la existencia de decenas y decenas de constructos latentes que del que

el lector acaba de imaginar. Supuesto que no esté ya inventado y disponga de su propio test en algún catálogo, el lector queda invitado a determinar cantidades de su constructo imaginario, y, a partir de ahí, a creer o hacer creer que de algún modo que tenga algún sentido, tal constructo absurdo también existe. La teoría clásica de tests, en cualquier caso, le ayudará a calcular su fiabilidad de diversas formas, y, si lo mide junto a un grupo de variados tests clásicos bien aceptados seguro que encuentra relaciones *interesantes* para analizar. Incluso podrá calcular intervalos de confianza en torno a sus puntuaciones “verdaderas”. Si lo hace, allá en Australia, James Lumsden se revolverá esta noche en su tumba.

La TCT es una vieja teoría, si es que es propiamente una teoría. Desde hace décadas la investigación especializada le presta poca atención, aunque por razones prácticas sigue siendo la más aplicada por los psicólogos y otros científicos sociales al construir, contrastar y utilizar instrumentos de medición.

Desde el punto de vista de la investigación y los grandes programas de medición, la TCT ya no es la teoría más popular. Susan Embretson describe el estado de cosas en estos términos: “Durante muchas décadas, los métodos de medición y los modelos psicométricos asociados permanecieron relativamente constantes. Según nos acercamos al final del siglo (XX), sin embargo, la medición y la teoría psicométrica está evolucionando a velocidad

creciente, debido primeramente a la tecnología informática. Los tests adaptativos computerizados en que los items son seleccionados individualmente y presentados para medir precisamente cada individuo, es el nuevo estado del arte. Sin embargo, la tecnología informática no permitiría por sí sola la disponibilidad de la medición adaptativa. La teoría psicométrica, a saber, la teoría de la respuesta al ítem (TRI), se ha desarrollado junto con la tecnología informática para permitir la selección de ítems adaptativa y la puntuación en la misma escala a través de infinitas formas de tests. La TRI es ahora el estado del arte en los métodos psicométricos.” (Embretson, 1999, pág. 407).

La TCT hace mucho que está bien integrada en el paisaje de la medición psicológica. En realidad, durante décadas no se podía hacer medición psicológica sin considerar al menos parcialmente la Teoría Clásica de Tests. En nuestro país todavía puede decirse que la inmensa mayoría de la medición psicológica práctica todavía se hace amparada, supuestamente, en la TCT. Creo que sería más exacto decir que en la actualidad la inmensa mayoría de la medición psicológica se hace bajo la coartada de la Teoría Clásica de Tests. Y con toda probabilidad la TCT seguirá siendo la teoría más utilizada en muchos trabajos aplicados y de investigación durante mucho tiempo por diversas razones prácticas. Entre ellas su idoneidad para problemas con muestras pequeñas, la accesibilidad de los cálculos o el hábito en los procedimientos científicos. Quiero resaltar, para acabar estas reflexiones, que una teoría de la validez

fundada sobre criterios sólidos, tal como aquí se expone al hablar de una concepción restrictiva de la validez, está a salvo de las dificultades que se han señalado para la teoría de la fiabilidad y de las dificultades inherentes a la inobservabilidad de los constructos. Esa teoría sólida de la validez no es más que la versión psicométrica del contraste de hipótesis que caracteriza a la ciencia empírica. Esa teoría sólida de la validez es suficiente para sostener los instrumentos de medición psicológicos y permitir juzgarlos por el grado en que son útiles para efectuar pronósticos útiles sobre la vida de las personas.