

Estatus Internacional Sobre el Desarrollo y Uso de test: Posibles Implicaciones para el Perú

César Merino-Soto ^{a1}, Thomas Oakland ^b, Sergio Domínguez-Lara ^c
& Anthony Córdova-Lonzoy ^d

Universidad de San Martín de Porres, Lima, Perú ^a; Universidad de Florida, Gainesville, USA ^b; Universidad Privada San Juan Bautista, Lima, Perú ^c; Universidad San Ignacio de Loyola, Lima, Perú ^d.

RESUMEN

El uso y desarrollo de tests se ha descrito como el buque de bandera de la psicología aplicada. Este trabajo discute el estatus internacional sobre el desarrollo y uso de los tests. Existen diferencias regionales e internacionales en el desarrollo y uso de los tests, y se explican en referencia a las condiciones necesarias para apoyar estos aspectos. La relevancia de los tests, como otras tecnologías, debería cambiar, a la luz de las necesidades cambiantes de quienes usan la información derivada de los tests. Se revisan algunos prominentes cambios a los cuales los tests deberían responder de acuerdo con el contexto de aplicabilidad. También, se exponen diversas condiciones externas e internas dentro de la psicología que impactan el uso y desarrollo de los tests. Estos y otros asuntos que impactan el desarrollo y uso de tests en Perú son discutidos.

Palabras Claves

Desarrollo de tests; uso de tests; evaluación; Comisión internacional de Tests; Perú.

ABSTRACT

Test development and use have been described as the flagship of applied psychology. This article discusses the status of test development and use internationally. Country and regional differences in test development and use exist and are explained considering the qualities needed to support test development and use. The relevance of tests, like other technologies, should change in light of changing needs of those who rely on test data. Some prominent changes to which tests should respond are reviewed. Several conditions that are external to and internal within psychology that impact test development and use are reviewed. These and other issues impact test development and use in Peru.

Keywords

Test development, test use, assessment, International Test Commission, Perú

¹ Correspondence about this article should be addressed to César Merino-Soto: sikayax@yahoo.com.ar

² **Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

INTERNATIONAL STATUS OF TEST DEVELOPMENT AND USE: POSSIBLE IMPLICATIONS FOR PERU

Las pruebas (o tests) se han descrito como el buque insignia de la psicología aplicada (Oakland, 2009); su aplicación en la mayoría de los países se da a lo largo de toda la vida de los seres humanos y puede ocurrir intensamente. Por ejemplo, a menudo ocurre primero al nacer, en que se suele utilizar una medida de Apgar como un método simple y confiable para evaluar rápida y sumariamente la salud de los recién nacidos. El uso de las pruebas puede extenderse a una evaluación de la calidad de vida junto con las cualidades cognitivas y físicas que a menudo se ven deterioradas en la senectud. Entre el nacimiento y la muerte, las pruebas se usan comúnmente para describir los comportamientos actuales, estimar comportamientos futuros, proporcionar orientación, ayudar a establecer intervenciones, evaluar el progreso, buscar necesidades especiales, diagnosticar, colocar o acreditar y retener o promover. El uso de las pruebas puede ser más prominente en las escuelas, con el omnipresente maestro que construye sus exámenes de fin de curso, y los exámenes de ingreso a la universidad que se utilizan comúnmente en la educación. El uso de pruebas también es común en la investigación y ha contribuido significativamente al establecimiento y crecimiento confiable y válido de la psicología y otras ciencias del comportamiento. La base empírica para la psicología es más fuerte en los países que tienen recursos para desarrollar pruebas de manera rigurosa (Oakland, Douglas, & Kane, 2016; Oakland, Wechsler, & Maree, 2013).

Un poco de historia

El primer uso registrado de pruebas se produjo en China hace aproximadamente 3000 años con el establecimiento de un examen del servicio civil de solución de problemas, percepción espacial visual, pensamiento divergente, creatividad y conocimiento de derechos y rituales (Zhang, 1988). La República Popular China continúa realizando exámenes nacionales cada primavera para ayudar a seleccionar empleados del gobierno. Por ejemplo, aproximadamente 16.000 solicitantes de las 1.600 posiciones del gobierno completaron el examen nacional de 2013 (Zhang, 1988).

Muchos psicólogos rastrean la evaluación moderna en los esfuerzos de tres psicólogos pioneros occidentales: William Wundt en Leipzig (Alemania), que estableció uno de los primeros laboratorios de psicología en 1879 para examinar la variabilidad individual en conciencia (i.e., experiencia interna; Escobar, 2016) desde los métodos de la fisiología; Galton en Londres, que estableció un laboratorio para estudiar las diferencias

individuales en un variado rango de conductas, aproximadamente diez años más tarde; y Alfred Binet en París, que estableció la primera prueba práctica de la inteligencia de un individuo, primero en 1904 y con revisiones en 1905 y 1911, una de ellas fue adaptada para uso en diversos países. El desarrollo de pruebas de inteligencia de grupo para uso en la Primera Guerra Mundial por parte de los Estados Unidos junto con la investigación sobre su confiabilidad y validez ayudó a establecer el respeto por su uso. El desarrollo de pruebas de aptitud para el uso en la Segunda Guerra Mundial por los Estados Unidos también contribuyó de manera importante a establecer el respeto por su uso (Oakland, 2004, 2012).

Estatus internacional del uso y desarrollo de tests

El uso de pruebas estandarizadas es una característica común de los países desarrollados. Dicho uso se produce en tres ámbitos: en las escuelas, las organizaciones y/o industrias y los servicios clínicos proporcionados por los profesionales. Prácticamente, las instituciones educativas tienden a ser el más grande consumidor de pruebas de un país, seguidas por las empresas y finalmente por los profesionales clínicos (Oakland, 2009; Oakland et al., 2013; Oakland et al., 2016).

Uso de pruebas con niños y jóvenes

Una encuesta sobre el desarrollo de pruebas psicológicas y el uso con niños y jóvenes en 44 países (Oakland, 1995, 2004) encontró que los psicólogos usaron medidas de inteligencia y personalidad con mayor frecuencia. Sorprendentemente, las medidas estandarizadas de aprovechamiento se utilizaron rara vez. Otros resultados de la encuesta encontraron que las Escalas de Inteligencia de Wechsler para los niños (WISC) era la prueba administrada individualmente de forma más frecuente. Además, las pruebas extranjeras desarrolladas se utilizaron más comúnmente que las pruebas creadas en el medio local, especialmente en los países más pequeños y menos industrializados. Habitualmente, las pruebas desarrolladas en el extranjero se tradujeron de manera simple al idioma local, y pocas fueron adecuadamente adaptadas para uso local. Además, la disponibilidad de normas nacionales representativas era algo rara. Las estimaciones de confiabilidad fueron más comunes que las evidencias de validez. Las medidas de la personalidad típicamente dependían de la teoría y carecían de normas. Los encuestados de casi todos los países reportaron la necesidad de pruebas adicionales para evaluar la discapacidad intelectual, las discapacidades de aprendizaje, el desajuste emocional y

social, el impedimento físico, los estudiantes talentosos y aquellos con impedimentos visuales o auditivos.

Uso de pruebas con adultos

Una encuesta sobre el desarrollo de pruebas y su uso en 29 países (Muñiz, Prieto, Almeida, & Bartram, 1999) también encontró que las pruebas desarrolladas en el extranjero eran más utilizadas que las desarrolladas localmente, con mayor frecuencia en países más pequeños y en vías de desarrollo. Entre ellos fueron los instrumentos más utilizados las Escalas de inteligencia de Wechsler para adultos (WAIS), el Indicador de Tipo Myers-Briggs, el MMPI y el NEO-PI.

Pruebas publicadas en español

El Buros Center for Testing de la University of Nebraska-Lincoln ha jugado un rol líder en facilitar el acceso a información sobre pruebas en idioma inglés durante aproximadamente 75 años. Sus Pruebas Publicadas en Español (Schlueter, Carlson, Geisinger, & Murphy, 2013) son una extensión del servicio, al proporcionar información valiosa que de otro modo es difícil de obtener, sobre 422 pruebas estandarizadas que fueron publicadas en español o traducidas al español. El propósito de las Pruebas Publicadas en Español es compilar y describir pruebas estandarizadas de lenguaje en español actualmente disponibles, en un esfuerzo para facilitar su apropiada selección por aquellos profesionales en la práctica clínica y de investigación, así como para brindar opciones evaluativas a psicólogos norteamericanos quienes frecuentemente evalúan usuarios hispanohablantes.

Sumario del estatus internacional del desarrollo y uso de Tests

De acuerdo a la literatura relevante (Muñiz et al., 1999; Oakland, 1995, 2004), el desarrollo y uso de las pruebas ha sido más fuertes en Australia, Canadá, la mayoría de los países de Europa Occidental y los Estados Unidos. Estos países muestran fuertes creencias en ciencia, tecnología, en las diferencias individuales y la meritocracia. El desarrollo y el uso de las pruebas aun parece ser escaso, pero emergente, en Asia, Europa del Este y América Latina. El desarrollo de las pruebas es más bajo en alrededor de los 22 países árabes y los 54 países africanos. Pocas pruebas están disponibles en 10 de los 12 países con las poblaciones más grandes (es decir, en orden de clasificación: República Popular China, India, Indonesia, Brasil, Pakistán, Bangladesh, Rusia, Japón, Filipinas y

Vietnam). Así, se podría estimar que un 80% de la población mundial reside en países que tienen ninguna o pocas pruebas psicológicas estandarizadas desarrolladas localmente.

Los esfuerzos peruanos para desarrollar y utilizar pruebas pueden beneficiarse de los recursos disponibles a través de TEA Ediciones, el mayor editor de pruebas de España (por ejemplo, con un catálogo de aproximadamente 300 pruebas), junto con El Manual Moderno (el mayor editor de pruebas de México; por ejemplo, con un catálogo de aproximadamente 200 Pruebas), así como Paidós (el mayor editor de pruebas de Argentina). Otros editores de pruebas están ampliando su cobertura internacional (por ejemplo, EOS de España, Pearson del Reino Unido, o Vetor, en Brasil).

Condiciones necesarias para el desarrollo de tests

La muestra estuvo conformada por 459 personas de Ecuador, que al momento se encuentran en modalidad de teletrabajo. De acuerdo con Israel (1992), esta muestra es representativa de grandes poblaciones. A muestra fue heterogénea, siendo que las personas no fueron de la misma organización. El 70% de la muestra indicó trabajar en el sector público, igualmente el 70% de la muestra hizo referencia a que formaba parte del sector de la educación. El 50% de la muestra fue de género masculino y el 96% indicó tener educación universitaria de pre y posgrado. El 92% indicó que no vivía solo, el 50% dijo que vivía con hijos menores y el 25% refirió que tiene un espacio privado para trabajar en la casa.

La necesidad de pruebas precede al desarrollo y uso de las pruebas

El desarrollo de pruebas se produce en respuesta a las necesidades profesionales y sociales y, por tanto, a sus mercados. Los profesionales que son potenciales consumidores de tests, deben ver la necesidad de tests y estar comprometidos con su compra y uso frecuente. Los consumidores de pruebas deben incluir a educadores, consejeros, especialistas en gestión, especialistas médicos, terapeutas ocupacionales, fisioterapeutas, psicólogos, trabajadores sociales, patólogos del habla y otros profesionales.

El público también constituye a los consumidores de tests, y deben conocer información sobre instrumentos válidos, confiables y cuyos resultados se mantengan confidencialmente. Esto incluye pagar servicios de evaluación que cumplan con lo anterior. Por otro lado, los profesionales que utilizan pruebas y aquellos cuyas vidas están

afectadas por el uso de las pruebas deben mostrar actitudes positivas hacia la ciencia y la tecnología con el fin de valorar el uso de la prueba (Carlson, & Geisinger, 2012).

Disponibilidad de instituciones educativas de apropiada preparación profesional

Los profesionales deben recibir educación adecuada y entrenamiento en los niveles de pregrado y posgrado para desarrollar y utilizar las pruebas apropiadamente. Por lo tanto, los programas universitarios de pregrado y posgrado son necesarios para preparar a especialistas en desarrollo y uso de pruebas. Los profesores de las cátedras relevantes suelen asumir el liderazgo para el desarrollo de pruebas o la investigación que examina sus cualidades psicométricas.

Esto requiere la disponibilidad de **profesores** con fuertes antecedentes académicos y profesionales en desarrollo y uso de pruebas, el cual es un requisito de recursos humanos que puede ser escaso en muchos países emergentes **como** Perú. Por ejemplo, pocos especialistas están preparados en psicometría, una especialización que se centra en el desarrollo de pruebas y su análisis. Los países pueden carecer de interés o recursos internos necesarios para fomentar el desarrollo y el uso racional de pruebas si a) se han basado en la teoría para orientar la psicología y han minimizado el valor de las intervenciones empíricas, b) no han mantenido el ritmo de los avances en estadísticas y psicometría, y c) su práctica ha descansado sobre algunas pruebas traducidas, pero ampliamente usadas, es decir, carentes de evidencias que acrediten su uso en el contexto peruano.

Disponibilidad de una infraestructura industrial de tests

Generalmente se necesita una industria de tests para ayudar en la adaptación, desarrollo, marketing y ventas de los tests. La base de esta industria puede ser pública (es decir, apoyada por el gobierno) o privada; pero en ambas situaciones, se deben emplear profesionales que muestren experiencia en las diversas especializaciones necesarias para desarrollar, adaptar y comercializar pruebas. El mercado de una prueba debe ser suficientemente grande para justificar su desarrollo, y el diseño de pruebas generalmente deben considerar su valor comercial. La adaptación o desarrollo de una prueba puede ser muy costosa, y aquellos que invierten su dinero, tiempo, talentos y otros recursos en su construcción deben esperar un retorno de sus inversiones. Por lo tanto, las pruebas deben ser compradas, no fotocopiadas. En resumen, una infraestructura debidamente financiada y administrada es necesaria para desarrollar y comercializar pruebas. Esto se halla

íntimamente vinculado al primer punto, que aborda la necesidad de pruebas. Es decir, si algunos psicólogos peruanos están acostumbrado a obtener versiones informales de las pruebas y aplicarlas sin evaluar sus propiedades métricas, y además esa acción no está regulada por el colegio profesional, ¿tendrá la necesidad de comprar una prueba creada en Perú cuyo costo probablemente sea 10 veces mayor al valor de la prueba informal cuyo uso no es cuestionado por los colegas, los entes profesionales, o el usuario? Probablemente no. Esto lleva al siguiente punto.

Disponibilidad de asociaciones profesionales relevantes

Se necesitan asociaciones profesionales para establecer y mantener altos estándares de desarrollo, adaptación y uso de pruebas. También son necesarios para promocionar su uso correcto. La fuerza de una profesión se ve directamente en la fuerza de sus asociaciones profesionales. Así, la fuerza de la profesión de psicología en el Perú puede verse en la fortaleza de su Colegio de Psicólogos del Perú (CPsP). El CPsP debe comprometerse firme y públicamente con las siguientes cuatro cualidades que apoyan su firme compromiso con el desarrollo y el uso de tests en psicología: a) valorar las diferencias individuales, b) orientarse cuantitativamente, c) abogar por el establecimiento y cumplimiento de altos estándares profesionales (por ejemplo, para el desarrollo, adaptación y uso de tests) y d) defender el valor del uso de pruebas para abordar importantes cuestiones sociales y personales. Este último criterio de calidad requiere de producción académica sobre el impacto del uso de las pruebas en Perú. En el año 2012, por iniciativa de un grupo de psicólogos interesados en el uso y aplicación de pruebas en Perú fue constituida la Comisión Nacional de Tests Psicológicos (CNTP), perteneciente al Colegio de Psicólogos del Perú (CPsP). Sin embargo, su actividad fue escasa, ya que existe registro que organizó un importante seminario internacional, y luego aparentemente estuvo inactiva, hasta su reactivación en septiembre de 2016. Se prevé que su rol será clave en un futuro, debido a las constantes regulaciones sobre el uso de pruebas y las exigencias de los organismos nacionales tanto públicos como privados. Respecto a ello, se puede plantear que queda claro que la sostenibilidad de esta orientación hacia las pruebas psicológicas debe ser independiente de los integrantes de la CNTP, y del directorio cambiante del CPsP, luego de cumplir con su periodo de gestión, dado que se trata de un interés gremial, no personal, y las repercusiones del (mal) uso de las pruebas cada vez toman mayor notoriedad.

Disponibilidad de estándares técnicos

El trabajo de personas dedicadas a desarrollar o adaptar pruebas y usarlas puede ser guiado por una línea académica que aborde estas cuestiones, escrita por profesionales que tienen una extensa experiencia académica y profesional. Las directrices y normas que afectan al desarrollo de pruebas son cuatro cuestiones separadas, pero entrelazadas: normas técnicas para el desarrollo de pruebas, normas éticas para su uso, aplicaciones clínicas de pruebas, y normas jurídicas que rigen el uso de pruebas. Esta toma de decisiones debería garantizar que la investigación realizada incluya toda la información debida, y para ello, diversas organizaciones profesionales han articulado normas en la presentación de estas evidencias en facetas (desde la conceptualización hasta el diseño y el análisis) que aborden la integridad de las medidas.

Por ejemplo, en 2013 la *European Federation of Psychologists Associations* aprobó normas que rigen el uso de las pruebas para la evaluación psicológica; estas normas de revisión de pruebas fueron aprobadas previamente (<http://www.ef.es/>). Por otro lado, la guía técnica más acreditada para el desarrollo y uso de las pruebas se proporciona en los *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association y National Council on Measurement in Education, 2014); adicionalmente, la guía técnica con mayor consenso internacional para la adaptación de las pruebas es proporcionada por las directrices la *International Test Commission* (www.intestcom.org).


La mayoría de los países no tienen asociaciones profesionales de psicólogos y por lo tanto no tienen códigos de ética establecidos a nivel nacional. En algunos países (Leach, & Oakland, 2007), los códigos de ética provienen de asociaciones profesionales que enfatizan cualidades amplias y virtuosas (por ejemplo, respeto, responsabilidad), que tienen un impacto generalizado en la práctica psicológica pero no se centran en intereses más específicos (es decir, práctica clínica, evaluación, etc.). Sólo el código ético de Letonia aborda explícitamente las cuestiones relativas a la adaptación de las pruebas. Por otro lado, el código de ética del profesional del psicólogo peruano (2004) aborda explícitamente las cuestiones relacionadas con la evaluación en su título VII, específicamente en el artículo 54 que indica que “*el psicólogo que tiene la responsabilidad de tomar decisiones sobre individuos, basándose en resultados de pruebas psicológicas (tests), tiene una adecuada comprensión de los problemas de medición, validez y confiabilidad.*” (p. 16)

Los *Ethical Principles of Psychologists and Code of Conduct* de la American Psychological Association de 2002 (www.apa.org/ethics/code.html) proporcionan la guía

más detallada para el uso ético de las pruebas. Por otro lado, *International Test Association's Guidelines for Test Use* discuten el uso justo y ético de las pruebas con la intención de proporcionar un marco acordado internacionalmente a partir del cual se puedan derivar las normas para la capacitación y la competencia de los usuarios de las pruebas (www.intestcom.org).

Las pautas clínicas para el uso de pruebas se encuentran comúnmente en artículos que se publican en revistas revisadas por pares, libros de texto autorizados y documentos de buenas prácticas de asociaciones profesionales. Por ejemplo, el *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, International Edition* (American Psychiatric Association, 2013), y el *International Classification of Diseases and Related Health Problems, Tenth Edition* (World Health Organization, 1992) generalmente guían el uso de datos extraídos de evaluaciones cuando se hacen diagnósticos. Las guías del *International Guidelines on Computer-Based and Internet-Delivered Testing*, del *International Test Commission*, sensibiliza a todas las partes interesadas en el proceso de evaluación, resaltando las directrices reconocidas internacionalmente que ponen de relieve las buenas prácticas en las tests, y especialmente en tests basadas en computadoras y/o realizadas a través de Internet (www.intestcom.org). Por otro lado, las normas legales que rigen el uso de pruebas, cuando están disponibles, generalmente preceden a las directrices técnicas, éticas y clínicas. El fundamento jurídico de los comportamientos suele definirse a nivel nacional, no a nivel internacional. Las leyes promulgadas, las jurisprudencias y las decisiones administrativas rigen la práctica de la psicología en todos los países, incluido el Perú. Se necesita investigaciones sobre este tema en el Perú.

La experiencia anecdótica con psicólogos peruanos generalmente confirma la ocurrencia de dos transgresiones ilegales y poco éticas más comunes relacionadas con el uso de la prueba: al igual que en otros países emergentes, plagio de test (es decir, tomar el producto de trabajo de alguien para beneficio personal sin compensar al autor) y no adherencia a las disposiciones del *copyright* (es decir, cuando las pruebas son fotocopias o adaptadas sin el consentimiento del autor de la prueba y editor) se practican comúnmente en Perú. Estas prácticas son contrarias a los esfuerzos para desarrollar pruebas en Perú y deben resultar en sanciones profesionales. Sin embargo, un argumento utilizado por los psicólogos para defender dicha práctica es que muchas veces el valor comercial de un test representa un tercio, o hasta la mitad del sueldo del psicólogo promedio peruano. Además, es conocido que las instituciones estatales no destinan el presupuesto suficiente para compras de pruebas originales. Por otro lado, las instituciones

privadas suelen dar libertad a los psicólogos para que usen sus propios instrumentos, y los seleccionen. Sin duda, son aspectos  reales del contexto peruano, y posiblemente latinoamericano, que deben tomarse en cuenta para comprender la conducta de adquisición de las pruebas psicológicas. También es bueno resaltar que, si existiera una consistente producción nacional de tests en Perú, los costos probablemente serían mucho más bajos en comparación a los provenientes de editoriales españolas o mexicanas. Esto ocurre, como ejemplo representativo, en Brasil con la editorial Casa do Psicólogo (<http://www.casadopsicologo.com.br/>), que además de comercializar pruebas de origen foráneo (p.e., WISC-IV), también lo hace con instrumentos creados en ese país, y los precios son relativamente menores. El tema de la comercialización nacional de pruebas propias, o extranjeras, inevitablemente viene con un aspecto que no se puede soslayar; es decir, los procedimientos realizados en los estudios instrumentales para establecer las propiedades psicométricas de las pruebas, forman parte de los esfuerzos del autor, quien invierte varios tipos de recursos, y sumado al proceso de producción por parte de la editorial, los costos implicados en su venta son una consecuencia esperada.

Cambios sociales y profesionales hacia los cuales el desarrollo y uso de pruebas deberían responder


Las prácticas de evaluación deben responder a las necesidades actuales dentro de un país, así como los de una profesión. Estas necesidades cambian y por lo tanto las prácticas de prueba necesitan cambiar. A continuación, se describen los cambios que se producen en varios países y pueden ser relevantes para el desarrollo y uso de las pruebas en el Perú.

Énfasis decreciente en las formas tradicionales de diagnóstico

En muchos países, la psicología clínica primero dependía de la psiquiatría y más tarde surgió de ella. Incluso, la evidencia anecdótica extraída de antiguos psicólogos clínicos puede señalar una época en que los psicólogos clínicos adquirieron en gran parte datos para ayudar a psiquiatras a llegar a sus diagnósticos. Los psicólogos clínicos generalmente mantuvieron el énfasis del modelo médico en el diagnóstico después de desarrollar una práctica independiente de la psiquiatría (Wampold, 2001; Wampold, Hollon, & Hill, 2011). El pensamiento predominante dentro de la psicología sugiere que el conocimiento del diagnóstico de una persona no identifica las intervenciones necesarias o los métodos para promover los comportamientos deseados (Wampold et al., 2011). En

este contexto, es plausible pensar que el público quiere y exige saber más que un diagnóstico.

Énfasis creciente en la descripción de la conducta dentro del contexto social y ambiental en que ocurre

Los métodos de evaluación deben adaptarse a este modelo, en el que el proceso de evaluación e interpretación de los datos de las pruebas se vea mejorado cuando se considera una actividad conjunta entre el profesional y aquellos con quienes el profesional está trabajando. Todo el mundo tiene una participación en el proceso de evaluación, incluyendo su interpretación y discusión sobre la naturaleza de las intervenciones deseadas, si es necesario, y los procesos para lograr los resultados deseados. Esto puede requerir la evaluación de diversos rasgos a través del uso del método de evaluación múltiple  para adquirir datos de diferentes fuentes confiables que describan comportamientos en múltiples contextos y períodos de tiempo.

La *World Health Organization's International Classification of Functioning, Disability and Health* (ICF; World Health Organization, 2001) proporciona un marco bio-psico-social para ver la conducta desde tres perspectivas diferentes: condiciones de salud, factores ambientales y factores personales. Todos interactúan, llevando al logro de rasgos latentes que tienen el potencial de ser exhibidos. El propósito de la ICF es describir la interacción compleja que varias cualidades ejercen sobre la salud, incluyendo la salud mental; por lo tanto, su propósito no es guiar los diagnósticos.

Énfasis en intervenciones que desarrollen o rehabiliten conductas

El público exige información que ayude a promover el desarrollo. El término *intervenciones respaldadas empíricamente* (a veces denominada como *basadas en evidencia*) se utiliza cada vez más en la literatura psicológica para enfatizar la necesidad de adquirir datos que informen sobre la naturaleza de las intervenciones, así como el proceso de mejora de los comportamientos deseados. El enfoque de la psicología en rasgos generales tales como inteligencia general, habilidades motoras finas y gruesas y comportamiento adaptativo es inconsistente con estos esfuerzos, pues no cumplen con criterios para intervenciones eficaces como consecuencia de un constante diálogo entre desarrollos que complementen el tratamiento y la valía de estos (Labrador, & Crespo, 2012). La eficacia aumenta cuando nos centramos en comportamientos pequeños y bien definidos, modificables, que se producen naturalmente en nuestro entorno; que nuestros

clientes y otros consideran importantes, y cuando la demostración de los comportamientos recién aprendidos se ve reforzada por uno mismo y por los demás. En resumen, los intentos de mejorar conceptos y rasgos generales (por ejemplo, inteligencia general) importantes para la psicología generalmente no han tenido éxito.

Adquisición de información funcional: vincular los datos a las intervenciones

Este aspecto puede ser deducido del párrafo anterior. Los psicólogos necesitan y desean probar los recursos que mejoran su capacidad de adquirir información funcional, en otras palabras, la que puede conducir a un impacto significativo en la vida de uno. Reconocen la necesidad de vincular los datos de evaluación con las intervenciones; por ejemplo, algunas pruebas desarrolladas en los Estados Unidos ofrecen sugerencias para formas de promover las destrezas deseadas de vida diaria que pueden estar poco desarrolladas o no ser bien utilizadas. Su uso permite a los usuarios de pruebas vincular los datos de evaluación con intervenciones empíricas. Por lo tanto, aquellos que participan en el desarrollo de pruebas y la evaluación deben tratar de identificar los comportamientos específicos que pueden ser modificados, que están influenciados por la edad (es decir, al cambio), y tienen valor funcional y clínico. Debe construirse herramientas informáticas que vinculen cada ítem de un test a intervenciones conductuales, de acuerdo con las evidencias de validez de los datos de la prueba para diversos fines. De este modo, los tests dejarían de ser considerados solo como elementos de apoyo al diagnóstico, sino también como una herramienta fundamental en la evaluación de intervenciones.

Mejorar la eficiencia del uso de los test

La práctica de la evaluación puede tomar muchas horas y por ende ser costosa. Como resultado, a los psicólogos se les pide que reduzcan la cantidad de tiempo necesario para recolectar, puntuar e informar sobre los resultados. Esto crea un dilema profesional: el psicólogo desea más información y merece ser compensado por su trabajo mientras el público puede ser incapaz de pagar, o no desea hacerlo. Las soluciones a este menudo dilema no son evidentes. Una forma de subsanar esta problemática incluye el uso de pruebas breves, las que pueden ser completadas sin una entrevista, ser administradas y anotadas por un técnico, y usando computadoras para adquirir, anotar e interpretar los datos.

Algunos defienden el uso de la teoría de la respuesta al ítem en lugar de métodos clásicos de evaluación al desarrollar y administrar pruebas cognitivas (Mpofu, &

Oakland, 2010). Estos métodos permiten a los desarrolladores de pruebas utilizar los datos de dificultad de los ítems para construir diferentes formas de prueba que sean apropiadas para cada examinado. Por lo tanto, aunque la prueba incluya ítems que difieren en dificultad, cada prueba es hecha a medida para incluir solamente aquellos ítems que son más discriminantes para la persona específica que se está evaluando.

Las decisiones relevantes requieren métodos de evaluación de alto nivel

Los resultados obtenidos de una prueba pueden tener diferentes impactos en la vida de una persona. Algunos resultados tienen un bajo impacto (por ejemplo, una puntuación en una prueba de ortografía); otros resultados tienen un alto impacto en las decisiones (por ejemplo, cuando una persona es admitida, contratada o despedida). Se necesitan diferentes estándares para el uso de pruebas en situaciones de bajo y alto impacto. Las pruebas de bajo impacto toleran menores niveles de confiabilidad y pueden depender solamente de evidencias basadas en el contenido, por ejemplo, basadas en entrevistas cognitivas (Howell, Phelps, Croft, Kirui, & Gitomer, 2013), en el análisis de la sustancialidad del constructo (Anderson & Gerbing, 1991; Castillo-Díaz & Padilla, 2013). Por ejemplo, las pruebas de rendimiento escolar informales, o las observaciones del rendimiento laboral, ambos para propósitos formativos. Por el contrario, las pruebas de más alto impacto requieren niveles más altos de consistencia interna (por ejemplo, coeficientes α superiores a .90 cuando se usan individualmente), así como evidencias de validez bien respaldadas, a fin de garantizar su interpretación y uso para las cuales fueron creadas, y minimizar el error inherente a todo proceso de medición. Estos esfuerzos pueden requerir la adquisición y análisis de datos clínicos considerables, y que pueden aumentar drásticamente los costos de construcción y desarrollo, y por lo tanto, justificar mayores precios de prueba. Las evaluaciones de alto impacto requieren generalmente la evaluación de rasgos múltiples a través del uso de un método de evaluación múltiple, para adquirir datos de diferentes fuentes confiables que describen comportamientos en contextos variados y en distintos períodos de tiempo. Estos procesos también incrementan los costos.

Recursos de evaluación son necesarios en países actualmente carentes de recursos necesarios para el desarrollo de tests

La disponibilidad de pruebas confiables y con sustantivas evidencias de validez en Perú aumentará en gran medida la competencia profesional para brindar servicios que

afectan una variedad de decisiones, incluyendo su uso para describir comportamientos actuales, estimar conductas futuras, proporcionar orientación, ayudar a establecer intervenciones, evaluar el progreso, diagnosticar, colocar y acreditar, retener / promover; e indirectamente, repercutirá a favor del prestigio del profesional psicólogo debido a que le proveerá herramientas con adecuadas evidencias de validez (International Test Commission, 2014). Las discusiones dentro del CPsP y entidades que se asocian para este fin, tendrán que considerar la relación costo-beneficio de la adaptación o la elaboración de nuevas pruebas. Aunque los que participan en estos debates inicialmente pueden preferir la elaboración de nuevas pruebas en lugar de adaptar pruebas bien establecidas, los modelos utilizados en otros países sugieren una mayor valoración en la adaptación inicial de las pruebas, al mismo tiempo de fortalecer la infraestructura de un país para asumir un papel más fuerte en el desarrollo de pruebas. Dos modelos que se han utilizado con éxito para adaptar las pruebas se resumen en otros lugares (Oakland, 2009).

En el primer modelo, los psicólogos de cuatro países trabajaron con el autor de una prueba, que identificó las mejores subpruebas para medir inteligencia, a la luz de la cultura objetivo, luego proporcionó los materiales visuales y capacitó al personal que recogió los datos de la norma en niños de cuatro países. El autor de la prueba ayudó en el análisis de datos, incluyendo la estructura de factores y tablas de normas. Utilizando el segundo modelo, una empresa de desarrollo de pruebas en un país de Europa del Este, obtuvo el permiso de las compañías de pruebas y de algunos autores de pruebas en los Estados Unidos, para adaptar algunas de las pruebas líderes del mundo, para su uso en su país. Dos o más pruebas se pueden adaptar simultáneamente, ahorrando costes y proporcionando datos de validez (Casillas & Robbins, 2005). Estos u otros modelos pueden ser relevantes para iniciar actividades de desarrollo de pruebas en Perú. Recientemente, una revisión de las prácticas, orientación de trabajo y panorama actual de la evaluación de la inteligencia en el Perú, ha revelado que hay aún retos que superar en este campo de evaluación (Blumen, 2016).

Condiciones externas e internas que pueden afectar el desarrollo y uso de pruebas en Perú

El desarrollo de la psicología, incluyendo sus características técnicas, se ve afectada por algunas condiciones externas a la propia ciencia, aquellas sobre las que la profesión tiene menos control, así como algunas que son internas a la misma psicología. A continuación, se analizan algunas condiciones externas e internas.


Algunas condiciones internas dentro de la psicología que impactan el desarrollo y uso de tests

El desarrollo y uso de test puede ser una actividad robusta cuando los psicólogos valoran las diferencias individuales y la meritocracia; valoran la ciencia, la tecnología y el empiricismo; se aseguran que los datos obtenidos de sus evaluaciones están salvaguardados y sean confidenciales; valoran los test como herramientas para identificar problemas sociales y personales importantes; y mantienen un alto estándar profesional en su trabajo.

Los profesionales necesitan comprar y usar pruebas, no fotocopiarlas; realizar investigaciones que examinen la confiabilidad, evidencias de validez y equidad de las pruebas; y para participar en los esfuerzos de desarrollo de pruebas que aborden importantes necesidades globales como autores y colaboradores, recopiladores de datos y revisores. Por ende, es importancia revisar y organizar la información peruana sobre aquellos atributos que impactan en la práctica de desarrollo, uso y adaptación de tests. Estos atributos hemos organizado en: a) las evidencias sobre las validaciones, b) las publicaciones normativas, c) las agrupaciones profesionales, y c) procesos en la construcción de tests:

Evidencias sobre validaciones

La investigación orientada psicométricamente realizada por estudiantes de pregrado y posgrado parece ser una actividad constante en las fases tempranas del desarrollo psicométrico peruano, pero rara vez ha se difundido ampliamente (Aliaga, & Giove, 1993, Lazo, 1980; Robles, 2013) y menos aún mediante su publicación en revistas científicas. Del mismo modo, en los últimos años en algunas universidades peruanas se han creado líneas de investigación psicométrica dirigida a estudiantes de pregrado con fines de titulación profesional, pero que sin un respaldo sólido a nivel conceptual, técnico y metodológico podría devenir en información poco relevante; e inclusive, esos instrumentos podrían ser utilizados en la práctica profesional por psicólogos sin el conocimiento suficiente como para evaluarlos de forma crítica, bajo el argumento de “ya está validado”, incurriendo en una falta ética, y perjudicando al evaluado,

A pesar de que los temas psicométricos rara vez aparec en las revistas peruanas durante la última década anterior (Livia, 2008), han sido publicados algunos trabajos sobre la validación de un selecto número de instrumentos destinados para evaluar la salud mental y los trastornos psiquiátricos en Perú (Pamo, Turín, & Pacheco, 1998; Anderson,

& Bravo, 2002; Valdivia, 1982; Varela, Ortíz, & Chávez, 2003). Esta información está incluida en algunas revistas peruanas representativas de las ciencias de la salud, como la Revista Médica Herediana y la Revista Herediana de Enfermería (Universidad Cayetano Heredia), la Revista de la Facultad de Medicina de la Universidad Nacional Mayor de San Marcos y la Revista Peruana de Salud Pública y Medicina Experimental (Instituto Nacional de Salud). Esta literatura discute típicamente la construcción y adaptación de pruebas, que son a menudo de autoinforme y medidas observacionales. Una revisión descriptiva y compensiosa sobre la tendencia actual peruana en el uso o reporte psicométrico habituales en Perú, aparentemente aún no se realizada desde la publicación de Livia (1990), y puede requerirse actualizar esta información.

En Perú a menudo las pruebas desarrolladas no reportan numerosos datos de confiabilidad y validez en sus manuales, o únicamente abordan aspectos parciales de las mismas (como la validez de la estructura interna, o la consistencia interna). Esto implica, que no se puede afirmar que tales resultados apoyan fuertemente su uso, y las afirmaciones sobre la validez que hacen los autores carecen de suficiente respaldo científico. Esto se confirma por la ausencia de publicaciones psicométricas sobre las mismas, y que conducen a afirmar que no cumplen con los estándares necesarios para su uso e interpretación (por ejemplo, Pereyra, 2013; Ruiz, 2003; Vicuña, 1998). Sin resultados psicométricos publicados que puedan ser evaluadas, no hay garantía que los instrumentos creados puedan ser defendibles para la práctica profesional y de investigación.

Hasta el momento, la investigación de las propiedades psicométricas de las pruebas psicológicas sólo podría mantenerse si las universidades desarrollan programas de investigación dirigidos al desarrollo y aplicación de métodos cuantitativos avanzados que fomenten la formación aplicada a la psicología cuantitativa, adquieren programas informáticos pertinentes, y mantienen un *staff* orientado hacia el desarrollo, validación y diseminación de conocimientos y procedimientos psicométricos, toda vez que estos van evolucionando a un ritmo acelerado. Esto también incluye la revisión evaluativa de las propiedades psicométricas de los instrumentos actualmente disponibles. A la fecha, apenas una revisión existe sobre un aspecto específico (e.g., la confiabilidad; Merino, Navarro, & García, 2014) sobre un instrumento de inteligencia emocional validado en el Perú (Ugarriza, & Pajares, 2005), pero se requieren evaluaciones realizadas consistentemente.

Publicaciones normativas

Es necesario en otra línea indicar que varias pruebas psicológicas recomendadas en guías normativas de uso clínico, laboral y jurídico no presentan la evidencia psicométrica suficiente para ser utilizadas para toma de decisiones en el ámbito profesional. Por ejemplo, en la *Guía de Valoración del Daño Psíquico en Personas Adultas Víctimas de Violencia Intencional* (Ministerio Público, 2016), usada en el ámbito de medicina legal, se sugiere el uso de varios de instrumentos antiguos. Por ejemplo, han sido sugeridos el BARSIT-Barranquilla (Olmo, 1958) o la primera edición del WAIS (Wechsler, 1955), entre otros; o se han indicado instrumentos que no cuentan con estudios psicométricos necesarios para el tipo de decisiones que rutinariamente se hacen en tales áreas. A su vez, en la *Guía de Práctica Clínica para el Diagnóstico y Tratamiento de la Psicosis en el Primer y Segundo Nivel de Atención* (Ministerio de Salud, 2008) se indica el uso del Cuestionario de Autorreporte de Síntomas (SRQ, por sus siglas en inglés; WHO, 1994) para decidir si una persona recibe tratamiento psiquiátrico; sin embargo, se declara como respaldo para su uso una tesis no publicada de maestría realizada hace más de 20 años (Véliz, 1996), y de la cual no hay mayor información recuperable. Además, no existen indicadores psicométricos actualizados que permitan valorar su relevancia para el propósito indicado, es decir, discriminar entre personas con y sin depresión clínicamente relevante. Este problema también se extiende a la amplia elección de algunos los instrumentos listados en un documento oficial del Instituto Nacional de Salud Mental (INSM, 2009), o en la versión actual de la misma (<http://www.insm.gob.pe/investigacion/basededatos.html>) en que se valora la validez y confiabilidad de los instrumentos. En esta recopilación la validez parece inducirse y justificarse racionalmente de la magnitud de los coeficientes alfa, y no fue evaluada al menos una presunción para elegir este coeficiente como indicador preciso de la confiabilidad. Lo apropiado de sugerir instrumentos para uso clínico sin un fuerte respaldo científico requiere ser nuevamente examinado en esos documentos.

Además, el INSM (2008) publicó un estudio documental de los instrumentos de evaluación psicológica creados, traducidos o validados entre los años 1990 y 2004 a nivel nacional con un total de 703 instrumentos. Sin embargo, dentro del informe figuran instrumentos desarrollados en el contexto de tesis de pre o posgrado, donde se puede observar pobre rigurosidad en el diseño y el análisis psicométrico, limitada discusión de las implicaciones psicométricas de los resultados y conclusiones superficiales sobre las



implicancias prácticas, éticas y teóricas. Otros estudios documentales han aparecido en la literatura de evaluación psicológica y psicometría peruana (Aliaga & Giove, 1993; Lazo, 1987; Valdivia, 1982), en que describía la instrumentación utilizada. En estos documentos, se puede identificar también que, comparado con la frecuencia de validaciones peruanas publicadas en revistas científicas, el mayor número de fuentes provienen de tesis; y que el objetivo esencial fue la descripción, pero no la revisión evaluativa de los trabajos seleccionados. En conjunto, los trabajos documentales realizados hasta la fecha parecen haber evaluado tangencialmente la cantidad y tipo de evidencias de validez de los instrumentos utilizados en estudios psicométricos, y que la nominación un estudio de como “psicométrico” (o vocablos derivados) parece ser la prueba suficiente de rigurosidad del estudio.

Agrupaciones profesionales

Además, una o más asociaciones profesionales solidas que abogan por el desarrollo y uso de las pruebas son necesarias, en parte, para establecer y hacer cumplir las normas técnicas, éticas y profesionales; promover la necesidad de programas de preparación para pregrado y posgrado; y ayudar a crear y publicar libros, revistas y otras publicaciones académicas que tratan temas de pruebas con el objetivo de hacer llegar esa información al público usuario. También necesitan trabajar en estrecha colaboración con las instituciones educativas para ayudar a asegurar que un número adecuado de profesionales estén siendo preparados y que su preparación sea consistente con las necesidades, normas y directrices actuales y emergentes.

Finalmente, existen algunas agrupaciones emergentes creadas para propulsar el desarrollo de test en el Perú, como la *Sociedad Peruana de Evaluación Psicológica* (<http://www.rppsicometria.com.pe/speps.htm>) y la *Sociedad Peruana de Psicometría*, creadas en el 2015; en los últimos años también han emergido grupos estudiantiles interesados en la evaluación y psicometría, iniciando capacitaciones propias. Aunque la fuerza de sus actividades no es bien conocida, son organizaciones nuevas y potenciales para estructurar actividades profesionales y académicas respecto a la evaluación psicológica. También se desconoce si hay comunalidad en las metas de estas organizaciones, por lo tanto, esta aparente descoordinación entre los equipos de trabajo emergentes, y la preferencia por la actuación individual, no permite la alineación conjunta de metas comunes, enfocadas hacia la ciencia psicológica y el colectivo profesional peruano. Instituciones como el *Instituto Brasileiro de Avaliação Psicológica* (IBAP;

www.ibapnet.org.br), o a nivel más general como la Asociación Iberoamericana de Evaluación Psicológica (AIDEP; www.aidep.org), pueden ser experiencias de agrupación útiles para el desarrollo local de este tipo de sociedades.

Construcción de instrumentos

Los psicólogos que se dedican a la psicología organizacional y social (León, 2013), educativa (Merino, Díaz, Zapata, & Benites, 2006) y clínica (Instituto Nacional de Salud Mental, 2008; Livia, 1990), generalmente usan test para evaluar atributos personales. Sin embargo, el conocimiento que necesitan estos psicólogos sobre la adaptación y desarrollo de pruebas psicológicas requiere una investigación local sobre las necesidades de evaluación, sus implicaciones científicas y sociales, y su publicación científica en revistas de alta calidad editorial. Un análisis situacional de esta cuestión en Perú puede llegar a concluir que estos conocimientos requieren fortalecerse, y ser vinculados con fuentes actualizadas en medición y evaluación psicológica. Asimismo, la deficiente preparación y educación en las estrategias de construcción y adaptación de test durante la formación de pregrado y postgrado, son límites relevantes para el uso informado de los tests, y el diseño de nuevas herramientas.

La literatura asociada a las características psicométricas de tests específicos, que pueden consultadas con el usuario, generalmente se encuentra en publicaciones de las revistas científicas. Esta debería ser la fuente principal de referencia bibliográfica para el investigador o profesional. Considerando estas fuentes, sin embargo, el número de artículos publicados que examinan las propiedades psicométricas con participantes peruanas, en revistas peruanas indizadas al menos en Scielo-Perú, es aún pequeño, llegando a un máximo de 3 anualmente en *Liberabit* (editada por la Universidad de San Martín de Porres) y en la *Revista de Psicología* (editada por la Pontificia Universidad Católica del Perú). Un aumento en la cantidad de esta información permitiría que los psicólogos y otras personas que usan pruebas puedan evaluar mejor la calidad de los instrumentos y realizar compras informadas e inteligentes. Sin embargo, los profesionales peruanos involucrados en el desarrollo o uso de pruebas que no tienen acceso personal o institucional a esta información académica tienen una oportunidad limitada de adquirir la información necesaria.

Considerando las publicaciones de propiedades psicométricas, se ha incrementado los estudios instrumentales en revistas peruanas de baja indización, donde no cuentan con estimaciones apropiadas de confiabilidad (e.g., Aiquipa, 2012; Gamero et al. 2016), ni

reportes de magnitud del efecto relevantes (e.g., García, 2016; Quijano, & Calderón, 2015) para respaldar sus afirmaciones de validez. Asimismo, algunos estudios presentan déficits cuantitativos y cualitativos en las evidencias de validez reportadas, que no permiten identificar con detalle los criterios utilizados para derivar las interpretaciones, y evaluar la evidencia necesaria que respalde el uso concreto del instrumento (e.g., Kaneko, 2013). Por último, y no menos importante, es que los baremos planteados (normalmente en percentiles) carecen de un reporte exhaustivo, de tal modo que existe una pobre identificación de la representatividad muestral, tamaño y adecuación del grupo normativo, así como controles técnicos que justifiquen la calidad del procesamiento analítico (por ejemplo, el piso y techo de los puntajes, tratamiento de puntajes extremos, suavización de frecuencias, etc.). Vinculado a este aspecto, la utilización de puntos de corte no es acompañada de argumentos racionales y/o criterios empíricos que avalen la validez de estas medidas para discriminar adecuadamente entre grupos (e.g., Ruiz, 2016; Silva, 2106). Estos estudios se enriquecerían mediante el vínculo con normativas internacionales sobre la adaptación de pruebas (e.g., International Test Commission, 2014), y la continuidad de su investigación psicométrica.

Sin lugar a duda, esto representa un reto para los profesionales implicados y para el CPsP, porque está implicado el uso de las pruebas, y sus consecuencias legales y éticas sobre los usuarios del test (el psicólogo profesional, la institución de referencia del psicólogo, y el cliente) y sobre la capacidad del CPsP para afrontar efectivamente estos problemas.

La actividad académica de los psicólogos peruanos relevantes para la evaluación psicológica puede encontrarse en revistas peruanas y no peruanas, pero el número de artículos publicados sobre validaciones para la población peruana en revistas científicas no es bien conocido aún. Una revisión anecdótica de las publicaciones al respecto puede resaltar el aumento de validaciones para la población, peruana, y algunos nuevos desarrollos psicométricos (por ejemplo, Bayes, Bazán, & García, 2012; Bazán, Bolfarine, & Branco, 2006) y estrategias de obtención de evidencias de validez (por ejemplo, Merino, 2016)

En diversas casas de estudio, han comenzado a incentivarse estudios psicométricos como conclusión del proceso formativo para la obtención del grado académico. Sin embargo, ante la falta de metodología actualizada aplicada en estos estudios, del control de calidad de los datos verificable, y de tribunales con fuerte experiencia en investigación psicométrica, sus resultados pueden contener un monto de

sesgo difícil de determinar. Con base al argumento anteriormente mencionado, los psicólogos aplicados en las distintas áreas aplicadas de la psicología peruana (e.g., educación especial, psicología de la salud ocupacional, etc.) probablemente requieran mejor especialización, o conocimientos avanzados sobre uso pertinente y ética en medición psicológica, para que consecuentemente se reduzca la utilización de pruebas suponiendo que las evidencias de validez obtenidas en otros contextos son también generalizables en el nuevo contexto de uso. Una implicancia de esta situación es que, plausiblemente, los psicólogos interesados e investigadores noveles puedan no valorar la experiencia y conocimientos especializados para estudiar psicométricamente una prueba, o mantener una actitud de rigurosidad científica a los procesos técnicos involucrados.

En el Perú, el desarrollo y uso inadecuado de pruebas no es solo un problema técnico o psicométrico, sino también ético respecto a su distribución y acceso. El problema se incrementa debido que estas mismas pruebas sin respaldo psicométrico también son comercializadas a los profesionales psicólogos, por los mismos autores o por agentes no asociados directamente con la psicología, dentro y fuera de algunas universidades. También es característico que puedan estar incluidas en las cátedras de enseñanza de pruebas psicológicas, o formar parte de la biblioteca de tests que generalmente están instaladas en la facultad de psicología en una universidad, sugiriendo que la evaluación de sus características no es el criterio principal para incluirlos. Aparentemente, el mayor desconocimiento de procedimientos en el desarrollo de test es influyente sobre el desconocimiento de las cualidades necesarias de un test para ser utilizado y evaluado por el profesional, y parece ocurrir más entre los psicólogos que residen fuera de Lima. Un ingrediente para este panorama son las características de contenido de la formación de pregrado y posgrado, respecto al tema de la evaluación y psicometría, un aspecto envolvente para todas las áreas no psicométricas (Prieto & Delgado, 2000). Por ejemplo, la actualización de su contenido mediante la incorporación de modernas concepciones de validez (e.g., ver Messick, 1980, 1994, 1995), vinculadas a enfoques psicométricos de modelamiento más elaborados comparado con la *Teoría Clásica de los Test* (e.g., *Teoría de Respuesta al Ítem*; Muñiz, 1997), requieren primero una revisión de su estado del arte, la posible unificación del currículo. Este tema específico no puede ser abordado en este manuscrito debido a su extensión y valor independientes del tema central del presente manuscrito.

En Conclusión

La Psicología en el Perú está preparada para entrar en una nueva fase de desarrollo, con el potencial de mejorar significativamente la investigación y el servicio desde el uso de tests psicológicos, siempre contando con el apoyo académico (e.g., universidades) y gremial (e.g., CPsP) para tal fin. La disponibilidad de tests bien desarrollados, que provean puntuaciones confiables e inferencias válidas, utilizados por académicos y profesionales bien preparados, conducirá a una mejor educación académica y profesional, y mejores servicios para niños, jóvenes y adultos en diferentes contextos.

Los procesos de organizaciones que sigan cabalmente estándares internacionales de evaluación y medición psicológica aún parecen en una fase emergente, sin embargo, es no implica la inexistencia de investigadores peruanos cuyos trabajos se encuentren ajustados a directrices actualizadas. Mediante una orientación técnica y teóricamente actualizada de los enfoques de validación, que implique también el desarrollo de modelos apropiados para el contexto, la profesión psicológica peruana tiene una oportunidad para ser valorada dentro de la vanguardia de la psicología académica y aplicada entre los países sudamericanos. Finalmente, aunque el contenido del presente manuscrito está orientado hacia la situación peruana, creemos que puede ser moderadamente generalizable para otros contextos latinoamericanos, y por lo tanto el presente aporte sirve como un marco de “lecciones aprendidas” para reflexión y modificación de la práctica.

Referencias

- Aiquipa, J. (2012). Diseño y validación del inventario de dependencia emocional – IDE. *Revista de Investigación en Psicología*, 15, 133-145.
<http://dx.doi.org/10.15381/rinvp.v15i1.3673>
- Aliaga, J., & Giove, A. (1993). *Baremos de tests psicológicos utilizados en el Perú*. Lima: Editorial Amauta.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Psychological Association.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders, fifth edition*. Washington DC: American Psychiatric Association.
- Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, 76(3), 732-740.
<https://doi.org/10.1037//0021-9010.76.5.732>
- Bayes, C., Bazán, J., & García, C. (2012). A new robust regression model for proportions. *Bayesian Analysis*, 7(4), 841-866. <https://doi.org/10.1214/12-ba728>
- Bazán, J., Bolfarine, H. & Branco, M. (2006). A skew item response model. *Bayesian Analysis*, 1(4), 861-892. <https://doi.org/10.1214/06-ba128>
- Blumen, S. (2016). New trends on intellectual assessment in Peru. *International Journal of School & Educational Psychology*, 4(4)254-261.
<https://doi.org/10.1080/21683603.2016.1166744>
- Carlson, J. F., & Geisinger, K. F. (2012). Test reviewing at the Buros Center for Testing. *International Journal of Testing*, 12(2), 122-135.
<https://doi.org/10.1080/15305058.2012.661003>
- Carrig, M., & Hoyle, R. (2011). Measurement Choices: Reliability, Validity, and Generalizability. En A.T. Panter & S. K Sterva (Eds), *Handbook of ethics in quantitative methodology* (pp. 127-157). New York: Taylor & Francis.
<https://doi.org/10.4324/9780203840023.ch5>
- Casillas, A., & Robbins, S. B. (2005). Test adaptation and cross-cultural assessment from a business perspective: Issues and recommendations. *International Journal of Testing*, 5(1), 5-21. https://doi.org/10.1207/s15327574ijt0501_2
- Castillo-Díaz, M. & Padilla, J. L. (2013). How cognitive interviewing can provide validity evidence of the response processes to scale items. *Social Indicators Research*, 114(3), 963-975. <https://doi.org/10.1007/s11205-012-0184-8>
- Colegio de Psicólogos del Perú (2004). Código de ética profesional. Recuperado de <http://www.colegiodepsicologosperu.org/w/imagenes/enlaces/archivos/CodigoEticaPeru.pdf>
- Escobar, R. (2016). El primer laboratorio de psicología experimental en México. *Revista Mexicana de Análisis de la Conducta*, 2(2), 116-144.
- Deacon, B. J. (2013). The biomedical model of mental disorder: A critical analysis of its validity, utility, and effects on psychotherapy research. *Clinical Psychology Review*, 33, 846-861. <http://dx.doi.org/10.1016/j.cpr.2012.09.007>
- Gamero, K., Flores, C., Arias, W. L., Ceballos, K. D., Román, A., & Marquina, E. (2016). Estandarización del Test de Dependencia al Celular para estudiantes universitarios de Arequipa. *Persona*, 19, 179-200.

- García, E. (2016). Propiedades psicométricas del cuestionario de dependencia emocional en estudiantes universitarios de la ciudad de Trujillo. *Cátedra Villarreal Psicología*, 1(1), 81-98.
- Howell, H., Phelps, G., Croft, A. J., Kirui, D., & Gitomer, D. (2013). *Cognitive interviews as a tool for investigating the validity of content knowledge for teaching assessments*. ETS Research Report Series, i-97. doi: <https://doi.org/10.1002/j.2333-8504.2013.tb02326.x>
- Instituto Nacional de Salud Mental (2008). Base de Datos de instrumentos de evaluación de Salud Mental y Psiquiatría. *Anales de Salud Mental*, 24(1). Recuperado de: <http://www.insm.gob.pe/investigacion/archivos/estudios/2008-ASM-BDIESMP/files/res/downloads/book.pdf>
- Instituto Nacional de Salud Mental (2009). Confiabilidad y validez de los cuestionarios de los establecimientos de salud de Lima y de la selva peruana. *Anales de Salud Mental*, 24(1). Recuperado de: <http://www.insm.gob.pe/investigacion/archivos/estudios/2009-ASM-CVCEESMLSP/files/res/downloads/book.pdf>
- Instituto Nacional de Salud Mental (2009). Confiabilidad y validez de los cuestionarios de los estudios epidemiológicos de salud mental de Lima y de la selva peruana. Lima: Autor. Disponible en: <http://www.insm.gob.pe/investigacion/archivos/estudios/2009-ASM-CVCEESMLSP/files/res/downloads/book.pdf>
- International Test Commission. (2014). ITC guidelines on quality control in scoring, test analysis, and reporting of test scores. *International Journal of Testing*, 14(3), 195-217. <https://doi.org/10.1080/15305058.2014.918040>
- Valdivia, O. (1982). *Bibliografía Psiquiátrica Peruana*. Lima: Litográfica del Perú.
- Kaneko, J. (2013). *Inventario de Personalidad con Tendencia a la Rotación Laboral Voluntaria Temprana*. Recuperado de: <https://pe.linkedin.com/in/juan-kaneko>
- Lazo, J. (1987). *Investigaciones Psicológicas en el Perú*. Lima: Gráfica Caribe.
- Leach, M. M., & Oakland, T. (2007). Ethics standards impacting test development and use: A review of 31 ethics codes impacting practices in 35 countries. *International Journal of Testing*, 7, 71-88. https://doi.org/10.1207/s15327574ijt0701_5
- León, F. (2013). Las psicologías del área social-organizacional en el Perú: 2003-2012. *Revista de Psicología*, 31(2), 177-226.
- Livia, J. (1990). Los tests psicológicos en el Perú, investigación, uso y abuso. *Revista Psicología actual*, 8(18), 23-32.
- Ugarriza, N., & Pajares, L. (2005). La evaluación de la inteligencia emocional a través del Inventario de Bar-On ICE: NA, en una muestra de niños y adolescentes. *Personas*, 8, 11-58.
- Livia, J. (2008). La producción científica y los estudios de post grado en psicología en el Perú. *Interamerican Journal of Psychology*, 42(3), 431-445.
- Messick, S. (1980). Test validity and ethics of assessment. *American Psychologist*, 35, 1012-1027. <https://doi.org/10.1037/0003-066X.35.11>
- Messick, S. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment*, 10, 1-9.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 15, 5-12. <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>

- Merino, C., & Allen, R. A. (2012). A factor-analytic study for the Bender Gestalt Test, 2nd edition: Internal structure and measurement model. *International Journal of Psychology*, 47 (ICP2012 supplement), 25.
- Merino, C., Díaz, M., Zapata, L., & Benites, L. (2006). School psychology in Peru. En S. R. Jimerson, T. O. Oakland & P. T. Farrell (Eds.), *The Handbook International of School Psychology*, (pp. 299 – 307). Oakland: Sage.
<https://doi.org/10.4135/9781412976138.n30>
- Merino, C., Pflucker, D., & Oakland, T. D. (2012). Temperamental styles in Peruvian children and adolescents: An exploratory and cross-cultural comparative study. *International Journal of Psychology*, 47 (ICP2012 supplement), 26
- Merino, C., Navarro, J., & García, W. (2014). Revisión de la consistencia interna del Inventario de Inteligencia Emocional de Bar-On, EQ-I: YV. *Revista Peruana de Psicología y Trabajo Social*, 3(1), 141-154.
- Merino-Soto, C. (2016). Percepción de la claridad de los ítems: Comparación del juicio de estudiantes y jueces expertos. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 14(2), 1469-1477.
<https://dx.doi.org/10.11600/1692715x.14239120615>
- Ministerio de Salud (2008). *Guía de práctica clínica para el diagnóstico y tratamiento de la psicosis en el primer y segundo nivel de atención*. Lima: Autor. Disponible en:
<http://www.minsa.gob.pe/dgsp/documentos/Guias/RM750-2008%20Psicosis%20I%20y%20II%20Nivel.pdf>
- Ministerio Público (2016). *Guía de valoración del daño psíquico en personas adultas víctimas de violencia intencional*. Lima: Autor. Disponible en:
<http://busquedas.elperuano.com.pe/download/url/anexo-de-la-res-n-3963-2016-mp-fn-que-aprobo-guias-elabora-anexo-res-n-3963-2016-mp-fn-1444631-1>
- Mpofu, E., & Oakland, T. (Eds.) (2010). *Assessment in Rehabilitation and Health*. Upper Saddle River, NJ: Merrill.
- Muñiz, J., Prieto, G., Almeida, L., & Bartram, D. (1999). Test use in Spain, Portugal and Latin American Countries. *European Journal of Psychological Assessment*, 15, 151-157. <https://doi.org/10.1027//1015-5759.15.2.151>
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- Oakland, T. (1995). Test use with children and youth internationally: Current status and future directions. In T. Oakland & R. Hambleton (Eds.), *International perspectives on academic assessment* (pp. 1-24). Boston, MA: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-0639-9_1
- Oakland, T. (2004). Use of educational and psychological tests internationally. *Applied Psychology: International Review*, 53, 157-172.
<https://doi.org/10.1111/j.1464-0597.2004.00166.x>
- Oakland, T. (2009). How universal are test development and use? In E. Grigorenko (Ed.), *Assessment of abilities and competencies in an era of globalization* (pp. 1-40). New York: Springer, Publisher.
- Oakland, T., Wechsler, S., & Maree, K. (2013). Test use with children across cultures: A view from three countries. In: Geisinger, K. F. (Ed.), *APA handbook of testing and assessment in psychology* (Vol. 3, pp. 231-258). Washington, DC: American Psychological Association.

- Oakland, T., Douglas, S., & Kane, H. (2016). Top ten standardized tests used internationally with children and youth by school psychologists in 64 countries: A 24-Year Follow-Up Study. *Journal of Psychoeducational Assessment*, 34(2), 166–176. <https://doi.org/10.1177/0734282915595303>
- Olmo, F. (1958). *Manual for the BARSIT-Barranquilla rapid survey intelligence test*. New York: The Psychological Corporation.
- Pamo, O., Turín, E., & Pacheco, G. (1998). Características de los artículos publicados en la Revista Médica Herediana, 1990-1996. *Revista Médica Herediana*, 9(1), 21-27. <https://doi.org/10.20453/rmh.v9i1.565>
- Pereyra, J. (2013). *Neotest*. Acceso el 16 de enero de 2014. Disponible en: <http://www.neotest.org/>
- Prieto, G., & Delgado, A. R. (2000). Utilidad y representación en la psicometría actual. *Metodología de las Ciencias del Comportamiento*, 2(2), 111-127.
- Quijano, J., & Calderón, S. (2015). Percepción de seguridad, estilos de afrontamiento y felicidad en adultos de centros poblados con alta y baja incidencia delictiva de la ciudad de Trujillo. *Revista de Psicología – UCV*, 17(1), 66-79. <https://doi.org/10.18050/revpsi.v17n1a6.2015>
- Ramos, L. M. S. (2016). Validez, confiabilidad y normas del cuestionario de autoconcepto físico en universitarios de Trujillo metropolitano. *Revista de Psicología*, 18(1), 9-25. <https://doi.org/10.18050/revpsi.v18n1a1.2016>
- Robles, Y. (2013). *Actualización de la base de datos de instrumentos de evaluación de salud mental y psiquiatría 2012. Informes Finales de Investigaciones*. Disponible en: <http://www.docenciaensaludmentalinsm.com/ojs-2.3.7/index.php/IFI/article/view/74>.
- Ruiz, A. (2005). Estandarización del test “Conociendo mis Emociones” de inteligencia emocional para escolares de 8 a 14 años. *Revista de Psicología – UCV*, 7, 155-124.
- Ruiz, C. (2003). *Test de autoestima para escolares*. Documento no publicado.
- Ruiz, P. (2016). Propiedades psicométricas del test de empatía cognitiva y afectiva en estudiantes no universitarios. *Cátedra Villarreal Psicología*, 1(1), 99-116
- Schlueter, J., Carlson, J., Geisinger, K., & Murphy, L. (2013). *Pruebas Publicadas en Español*: Lincoln, Nebraska: The University of Nebraska-Lincoln.
- Sogi, C., Perales, A., Anderson, A., & Bravo, E. (2002). Producción científica de los investigadores de la Facultad de Medicina, UNMSM: Tendencia 1991-2000. *Anales de la Facultad de Medicina*, 3, 191-200. <https://doi.org/10.15381/anales.v63i3.1499>
- Valdivia, O. (1982). *Bibliografía Psiquiátrica Peruana*. Lima: Litográfica del Perú.
- Varela, L., Ortiz, P., & Chávez, H. (2003). Características de los trabajos científicos nacionales en geriatría y gerontología 1980-2001. *Revista Médica Herediana*, 14(1), 18-25. <https://doi.org/10.20453/rmh.v14i1.767>
- Vicuña, L. (1998). *Inventario de Interés Vocacionales y Profesionales: CASM 84-R-98*. Lima: CEDEIS.
- Vicuña, L. (2002). Elaboración del Inventario de Intereses Vocacionales para carreras Técnicas y de Oficio. *Revista de Investigación en Psicología*, 5, 117-140. <https://doi.org/10.15381/rinvp.v5i1.5060>
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. New York: Routledge.
- Wampold, B. E., Hollon, S. D., & Hill, C. E. (2011). Unresolved questions and future directions in psychotherapy research. In: J. C. Norcross, G. R.

- VandenBos, & D. K. Freedheim (Eds.), *History of psychotherapy: Continuity and change* (2nd ed.). Washington, DC: American Psychological Association.
- Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale (WAIS)*. New York, NY: Psychological Corporation
- World Health Organization (1992). *International statistical classification of diseases and related health problems, tenth revision (ICD-10)*. Geneva, Switzerland, Author.
- World Health Organization, (1994). *A user's guide to the Self Reporting Questionnaire (SRQ)*. Geneva, Switzerland: Author.
- World Health Organization (2001). *International classification of functioning, disability and health (ICF)*. Geneva, Switzerland: Author.
- Zhang, H. (1988). Psychological measurement in China. *International Journal of Psychology*, 23, 101-117. <https://doi.org/10.1080/00207598808247755>.

Received: 2018-05-18

Accepted: 2021-02-26