
Aproximació històrica i conceptes bàsics de la psicometria

PID_00216191

Julio Meneses

Temps mínim de dedicació recomanat: 4 hores



**Julio Meneses**

Professor dels Estudis de Psicologia i Ciències de l'Educació de la Universitat Oberta de Catalunya (UOC) i investigador de l'Internet Interdisciplinary Institute (IN3). Llicenciat en Psicologia per la Universitat d'Oviedo, es va doctorar en el programa de Societat de la Informació i el Coneixement de la UOC i, actualment, treballa com a responsable de la metodologia en diversos projectes d'investigació a les ciències socials. La seva recerca s'orienta a l'aplicació dels mètodes de recerca quantitativa i a l'ús de tècniques d'anàlisi multivariant, particularment en les àrees de la introducció de les TIC als sistemes educatius, la relació dels infants i els joves amb la tecnologia i l'abandonament i el reingrés dels estudiants universitaris.

Índex

| | |
|--|-----------|
| Introducció..... | 5 |
| 1. La psicometria en el context de la psicologia..... | 7 |
| 1.1. Una aproximació històrica a la psicometria | 7 |
| 1.2. La psicometria avui | 12 |
| 1.3. La psicometria en el context de l'avaluació psicològica | 15 |
| 2. Fonaments de la psicometria..... | 18 |
| 2.1. Definició i classificació dels tests | 18 |
| 2.2. Models de mesura psicomètrica | 22 |
| 2.3. Teoria clàssica dels tests | 26 |
| 2.4. El procés d'inferència psicomètrica | 29 |
| 3. Construcció i administració de tests..... | 33 |
| 3.1. El procés de construcció de tests | 33 |
| 3.2. Criteris per a la valoració de tests | 36 |
| 3.3. Aspectes ètics i deontològics en l'ús de tests | 39 |
| Bibliografia..... | 47 |

Introducció

Aquesta introducció a la psicometria es proposa donar a l'estudiant algunes claus importants per a abordar la complexitat dels coneixements, procediments i valors vinculats amb el desenvolupament i l'administració de tests. Partint de l'enquadrament de la psicometria en el context general de la psicologia, començarem desenvolupant una aproximació històrica al seu naixement, des dels antecedents remots fins als primers desenvolupaments, descriurem l'estatus actual de la psicometria com a disciplina científica i, finalment, construirem una definició formal tenint en compte el paper que tenen els tests com a instruments d'avaluació psicològica. A continuació abordarem els fonaments de la psicometria, en què tindrem l'oportunitat d'oferir una definició i classificació dels tests, abordar els models de mesura psicomètrica, introduir la teoria clàssica dels tests i recapitular el procés d'inferència psicomètrica en què es basen els tests. Finalment, discutirem les qüestions relatives a la construcció i administració de tests, tractant les diferents fases implicades en el desenvolupament i oferint alguns criteris importants per a la valoració dels tests disponibles en la literatura. Conclourem aquesta introducció amb una discussió dels aspectes ètics i deontològics vinculats amb la utilització de tests en la pràctica professional de la psicologia.

1. La psicometria en el context de la psicologia

Entendre un camp d'estudi determinat implica, en la majoria dels casos, conèixer-ne les arrels fundacionals. És a dir, conèixer els problemes a què es van enfrontar els pioners, la manera en què els entenien i les solucions que hi van donar. Tot i que aquest no és el lloc per a fer un viatge en profunditat¹, una aproximació històrica a la psicometria ha de començar recollint el que es consideren els seus orígens i antecedents. Malgrat que el desenvolupament i l'administració de tests psicològics és una pràctica desenvolupada fonamentalment a partir del segle XX, és possible trobar alguns antecedents remots en cultures tan antigues com la xinesa. Partint d'aquests antecedents, i repassant algunes de les contribucions més importants al seu desenvolupament com a disciplina científica, estarem en disposició d'oferir una definició formal de la psicometria que ens permeti situar-la en el context general de la psicologia.

⁽¹⁾Entre d'altres, vegeu els treballs clàssics de Goodenough (1949), DuBois (1970), i Hilgard (1987), més recentment, Buchanan i Finch (2005), o Jones i Thissen (2007).

1.1. Una aproximació històrica a la psicometria

Tal com s'acostuma a assenyalar, el desenvolupament de les primeres dinasties de l'antic imperi xinès va generar els primers sistemes d'avaluació dels individus en funció de la seva habilitat. Tot i que alguns cronistes han apuntat a referències tan antigues com les de l'any 2000 aC², l'estudi de les evidències arqueològiques ha posat en dubte aquesta antiguitat. En qualsevol cas, tal com ha apuntat Bowman (1989), les proves documentals que es conserven permeten situar aquest origen en un període relativament més recent, durant la dinastia Tang (618-907). Durant aquests anys es va desenvolupar un sistema d'avaluació imperial que va permetre la selecció i promoció dels funcionaris dels diferents departaments de l'Administració. Aquest sistema va tenir un important impuls durant la dinastia Ming (1368-1644) quan es va establir un examen institucional segons el mèrit per a tots els funcionaris dels diferents nivells territorials –des del nivell municipal al nacional–, posant en marxa un dels primers sistemes de classificació oficial mitjançant l'expedició dels primers títols formals que acreditaven l'accés als diferents nivells de responsabilitat.

⁽²⁾Per exemple, podeu veure DuBois (1970).

Aquests sistemes d'avaluació són fàcilment homologables als que l'educació formal europea, especialment la universitària, va desenvolupar a partir de la introducció dels exàmens orals per als seus estudiants a partir del segle XIII. Tal com assenyala Rogers (1995), la invenció i la incorporació del paper en la vida quotidiana a partir del segle XVI va facilitar el trànsit cap a les proves escrites que durant el segle XIX van esdevenir els primers mecanismes de selecció competitiva dels estudiants universitaris. Com a antecedents remots de la psicometria, tant l'avaluació en l'educació formal com l'establiment del sistema d'avaluació imperial van portar un canvi important en la concepció sobre el

judici de les capacitats de les persones. D'aquesta manera, es va anar traslladant la confiança en el judici personal basat en impressions cap a l'administració de proves institucionals basades en una autoritat imparcial que objectivava les destreses i coneixements requerits en els àmbits educatiu i administratiu.

Tot i que algunes vegades s'ha passat per alt, un altre antecedent remot es trobaria a l'inici de l'**avaluació psiquiàtrica** a mitjan segle XIX. D'acord amb el treball de Bondy (1974), és convenient tenir present l'esforç dels primers professionals orientats a l'estudi dels problemes mentals i les lesions cerebrals en l'establiment del que es podrien considerar les primeres proves d'avaluació psicològica. Així, per exemple, es van desenvolupar els primers tests per a avaluar les conseqüències del dany cerebral, algun d'ells tan elaborat que exigia dedicar-hi períodes de 100 hores. Malgrat les seves limitacions, entre d'altres per l'absència de procediments estàndards en el seu ús, aquestes primeres proves van incorporar molts dels elements que, amb l'evolució de l'estudi dels símptomes psicològics, encara són utilitzats en les proves de diagnosi actuals.

Desenvolupades les bases per a l'examen individual, tant en contextos d'activitat quotidiana com en situacions de trastorn psicològic o accident, els antecedents recents de la psicometria es trobarien en el desenvolupament de l'**estudi sistemàtic de les diferències humanes** durant el segle XIX. Primer gràcies als treballs de Friedrich W. Bessel (1784-1846) i Carl F. Gauss (1777-1855) –que van ser pioners en l'estudi de les diferències individuals en la percepció en el camp de l'astronomia– i després a les contribucions de Gustav T. Fechner (1801-1887) i Hermann von Helmholtz (1821-1894) en el desenvolupament de la psicofísica –que va representar l'inici de la psicologia com a disciplina acadèmica–, totes dues aproximacions van constituir un avenç important en la sistematització de la mesura de les sensacions psicològiques produïdes per l'estimulació física (Boring, 1978, per a una revisió històrica). Així, mitjançant el desenvolupament de les primeres lleis que relacionen estímuls físics i sensacions, es van anar assentant les bases per a la mesura psicològica. D'aquesta manera, es van perfeccionar els mètodes de presentació dels estímuls i el registre de les respostes, es va treballar en la millora de la precisió de les mesures, i es van adoptar condicions controlades per a la seva consecució. Tots ells elements indispensables per al plantejament posterior dels problemes més importants que la psicometria hauria d'afrontar en el seu desenvolupament inicial.

Així, una vegada apuntats els antecedents remots i més recents, hem d'esmentar els treballs de Sir Francis Galton (1822-1911), James McKeen Cattell (1860-1944) i Alfred Binet (1857-1911) com els veritables **pioners de la psicometria moderna**. Com a continuadors de les aproximacions anteriors a l'estudi sistemàtic de les diferències humanes, Galton i Cattell van contribuir necessàriament a l'establiment de la psicologia experimental a Europa i els Estats Units, respectivament, mitjançant la creació dels primers laboratoris antropomètrics per a l'estudi de les característiques humanes (Valentine, 1999, i Sokal, 1982) i el desenvolupament dels primers tests per a l'avaluació de les diferències sensorials, perceptives i de comportament. Obrint un nou camí per

a l'estudi científic de la psicologia, tots dos autors es van aventurar a postular relacions entre les seves mesures i l'intel·lecte, i van arribar a conclusions a vegades controvertides i d'altres àmpliament criticades amb posterioritat pel seu simplisme. Per la seva banda, Binet va adoptar un enfocament innovador i va ser responsable del que es considera el primer test d'aplicació general per a la mesura de les habilitats cognitives.

Responent a la necessitat d'identificar els estudiants amb problemes per a assolir els objectius de la instrucció ordinària a les escoles de París del principi del segle XX, Binet i el seu col·lega Théodore Simon van recollir els incipients moviments en l'avaluació de la discapacitat cognitiva mitjançant bateries de tests (Nicolas i Ferrand, 2002). A partir d'aquests, i a petició del Ministeri d'Educació francès, el 1905 van desenvolupar la primera prova que va permetre classificar els infants segons la intel·ligència (Wolf, 1969). Aquest treball va permetre el naixement de l'interès de la psicologia per l'atenció a les necessitats educatives especials, però no va ser fins al 1908 quan aquests autors van publicar una revisió de la seva escala que va permetre mesurar el que es va anomenar *el nivell mental*. Gràcies a l'adopció d'un grup de referència compost per nens que tenien entre tres i tretze anys, i una vegada ordenats els ítems en funció de l'edat en què eren típicament resolts, van desenvolupar el primer test per a quantificar la intel·ligència. Aquesta quantificació, prenent com a referència l'edat biològica de l'infant, va conduir a la definició del que es coneix, fins a l'actualitat, com *l'edat mental*.

Adoptant el mateix enfocament orientat a la pràctica, la incipient psicometria es va veure impulsada en els circuits acadèmics entorn del debat sobre la **mesura de la intel·ligència** mitjançant el desenvolupament de tests (Martin, 1997). A partir dels treballs de Binet en el camp de l'educació, l'interès per l'avaluació de la cognició va fer el salt al circuit internacional de la mà de Lewis M. Terman (1877-1956), el qual va publicar el 1916 la revisió Stanford-Binet. D'acord amb el relat de Goodenough (1949), aquesta contribució va establir una fita important en el desenvolupament dels tests tal com els coneixem actualment. Revisant els ítems originals, incorporant pautes clares d'administració i destacant la importància de la representativitat de les mostres per a la interpretació correcta de les puntuacions, va esdevenir l'escala de referència per a la mesura de la intel·ligència durant les dècades següents. De fet, va fer un altre pas important per al desenvolupament de la psicometria, en la mesura en què es va estendre l'ús d'aquesta escala a l'exèrcit nord-americà durant la Primera Guerra Mundial. En aquest context, i gràcies al suport del Govern dels Estats Units, la psicometria va créixer en el centre de les polítiques militars³.

⁽³⁾Entre d'altres, podeu veure Zeidner i Drucker (1988).

Recuperant l'esperit classificador, els tests haurien de servir com a instruments d'avaluació psicològica per al reclutament dels nous soldats. Així, a partir de la generalització de l'ús de la revisió Stanford-Binet, es van desenvolupar nous tests i es va desenvolupar l'administració en grup per fer més eficient el procediment de mesura. D'aquests esforços, entre d'altres, cal destacar l'empremta de Robert Yerkes (1876-1956) en el desenvolupament dels tests Army Alpha i

Beta (Carson, 1993), que també van abordar una controvèrsia important entorn dels **biaixos culturals dels tests** que es va anar construint durant aquestes primeres dècades del segle XX. Tal com es va discutir àmpliament (Jensen, 1980), les noves proves per a mesurar la intel·ligència podien no estar lliures d'influències culturals de manera que subestimarien els que no parlessin la llengua anglesa, els analfabets i qui patís alguna discapacitat visual o auditiva. Recollint els incipients moviments entorn de les proves no verbals, la versió beta del test d'intel·ligència de l'exèrcit nord-americà va representar el reconeixement de la importància d'aquestes diferències, com també la necessitat de minimitzar-les en qualsevol context en què s'utilitzin els tests com a instruments d'avaluació psicològica.

Una qüestió interessant, per les conseqüències que va tenir més endavant en el desenvolupament de la psicometria, va ser la concepció que es va anar construint entorn del mateix concepte d'*intel·ligència*. Arrelada a les nocions biològiques i hereditàries dels pioners, i sustentada en els desenvolupaments estadístics de Charles Spearman (1863-1945), un corrent important va concebre la intel·ligència com un únic factor –el factor *g*, segons Spearman– que explicaria les puntuacions en els diferents tests d'intel·ligència desenvolupats fins aleshores. En canvi, des de la posició “alternativa” que va encapçalar Louis Leon Thurstone (1887-1955), la intel·ligència estaria composta en realitat per diversos factors específics. Els seus avenços en el desenvolupament estadístic de les **tècniques d'anàlisi factorial** que va introduir Spearman van permetre comprendre millor aquesta aparent contradicció⁴. De fet, es podia entendre el factor general d'intel·ligència com una mena de factor subjacent que recolliria la variabilitat de les aptituds específiques de Thurstone, tant si eren relatives a la comprensió verbal com a la competència numèrica o a l'habilitat espacial, entre d'altres. De la mateixa manera, una vegada controlat l'efecte del factor general, el plantejament d'uns factors grupals per part de Spearman permetria recollir la variabilitat comuna observada en tests que comparteixen demandes verbals, numèriques o espacials similars. Més enllà de les conseqüències teòriques sobre la mesura de la intel·ligència, el perfeccionament de l'anàlisi factorial va permetre pavimentar el camí estadístic per al desenvolupament de la psicometria fins al final del segle XX.

⁽⁴⁾Per exemple, podeu veure Cattell (1943).

A més a més, aquesta visió de la intel·ligència com un conjunt determinat d'aptituds cognitives específiques va contribuir necessàriament al desenvolupament dels **primers estudis sistemàtics de validesa**. Tal com va passar durant la Primera Guerra Mundial, el Govern nord-americà va tornar a incorporar aquestes pràctiques als seus programes de reclutament uns quants anys abans de participar en la Segona Guerra Mundial. Tot i que els esforços anteriors havien donat els seus fruits, l'exèrcit nord-americà va trobar en aquest plantejament la manera de resoldre algunes limitacions de l'enfocament unitari de la intel·ligència, principalment per la seva escassa capacitat per a seleccionar candidats amb perfils molt especialitzats. Va ser durant aquest període quan, sota la direcció de John C. Flanagan (1906-1996), les forces d'aviació van administrar un conjunt de bateries de tests per a seleccionar i classificar

els pilots, enginyers de vol i altres tècnics encarregats dels instruments de navegació. L'estudi de la relació entre les puntuacions en les diferents aptituds i l'èxit en la formació posterior dels reclutes va ser clau en aquest procés (Goslin, 1963), permetent la millora dels procediments de selecció i classificació, i conformant un dels primers estudis empírics de la validesa de les mesures psicològiques.

No hem d'oblidar, però, el treball que paral·lelament es va a portat a terme amb relació a l'**avaluació de la personalitat**. Encara en el context de les polítiques militars de la Primera Guerra Mundial, el Govern dels Estats Units es va enfrontar a un altre tipus de problema pràctic. Més enllà de la selecció i classificació dels reclutes, un important esforç es va adreçar a la identificació de candidats susceptibles de patir trastorns psicològics. Robert S. Woodworth (1869-1962) va ser l'encarregat de desenvolupar un nou tipus de prova que permetés avaluar l'estabilitat emocional dels soldats per a minimitzar-ne la presència a les seves files. Aquest test, el seu *Personal Data Sheet*, va introduir un conjunt de preguntes amb resposta positiva o negativa que, a diferència de l'avaluació de la intel·ligència, no contenia respostes necessàriament correctes o incorrectes. D'aquesta manera, Woodworth va elaborar una prova que permetia detectar els casos problemàtics, no ja amb la comparació de la resposta individual amb una mostra representativa de la població, sinó a partir de les respostes donades per individus amb trastorns psicològics ja diagnosticats. Aquest treball va assentar les bases per al desenvolupament dels tests d'avaluació de la personalitat posteriors (Gibby i Zickar, 2008), com també va començar a definir les bases per a tractar adequadament la possibilitat de frau en la resposta, tant per a amagar com per a fer veure que es pateix un trastorn.

El prolífic Thurstone, que va desenvolupar les primeres proves d'anàlisi de consistència interna dels tests en l'àmbit de l'avaluació de la personalitat, va fer una altra contribució important al desenvolupament de la psicometria. Partint dels experiments desenvolupats pels pioners (Gulliksen, 1968), especialment en la branca de la psicofísica, el 1927 va proposar la llei del judici comparatiu com el mètode per a la **mesura de les actituds, les preferències i els valors**. Aquesta innovació va consistir a traslladar el judici perceptiu sobre parelles d'estímul físics –per exemple, la tria entre dos sons segons la intensitat– a la valoració de característiques psicològiques no estrictament relacionades amb les propietats físiques –per exemple, la tria entre dos comportaments segons l'acceptabilitat. Després de proposar una segona llei, la llei del judici categòric, l'aproximació de Thurstone a la mesura de les actituds estava preparada per a ser aplicada al judici general, és a dir, sense requerir la comparació entre parelles de fenòmens. D'aquesta manera, tal com pretenien Thurstone i els seus continuadors (Bock i Jones, 1968), la psicometria disposaria d'un mètode d'escalament per a mesurar i tractar numèricament valoracions individuals subjectives en la cerca d'una representació objectiva dels fenòmens psicològics. Aquest mètode va donar un suport metodològic important a les dècades següents de recerca en psicologia social.

1.2. La psicometria avui

Tot i la dificultat de fer una aproximació històrica a partir de la selecció d'algunes de les aportacions més importants, aquest exercici ens permet il·lustrar el naixement d'una disciplina a partir dels problemes i les diferents aproximacions dels pioners. La psicometria, un nou espai de treball metodològic entorn del desenvolupament i l'administració de tests, es va anar conformant amb un component aplicat, orientada a les demandes en diferents contextos, i va formar part d'alguns dels debats teòrics més importants de la mateixa psicologia. No és, però, fins a la dècada dels anys trenta del segle XX que podem situar l'**inici de la seva constitució com a disciplina científica** tal com la coneixem actualment. És important començar assenyalant de nou el paper decisiu de Thurstone, el qual, amb l'objectiu d'establir i promoure la psicologia com una ciència quantitativa (Samejima, 2000), el 1935 va fundar la Psychometric Society i en va ser el primer president. A més a més, va ser l'impulsor de la primera revista especialitzada que encara és de referència obligada, la revista *Psychometrika*, que va anar publicant alguns dels treballs més importants sobre els quals es va formalitzar la psicometria. Pocs anys després, el 1946, va ser també el primer president de la Divisió d'Avaluació i Mesura de l'American Psychological Association.

En aquest sentit, podríem definir el període entre les dècades dels anys trenta i els seixanta com l'**època daurada de la psicometria**. És en aquest període que es publiquen, a més a més, els llibres i manuals més importants que la vertebrarien. Entre els manuals cal destacar *The reliability and validity of tests*, de Thurstone (1931), que sistematitzava el que s'havia desenvolupat fins aleshores amb relació a la teoria dels tests i suggeria el rol central de la fiabilitat com a requisit per a la validesa de les mesures en la psicometria. Tornarem més endavant sobre aquestes qüestions, elements bàsics de la mesura indirecta dels fenòmens psicològics mitjançant tests. En la mateixa època, també va ser important la primera edició del manual *Psychometric methods*, de Guilford (1936), un intent per organitzar el camp popi de la psicometria entorn de la teoria dels tests, l'escalament psicològic i el psicofísic. La teoria clàssica dels tests, com va ser àmpliament coneguda a partir dels treballs de Spearman sobre l'estimació dels errors de mesura, començava el seu camí en els circuits docents universitaris encarregats de formar els futurs psicòlegs.

D'altra banda, durant els anys trenta i especialment els quaranta, un dels debats més apassionants oberts per la psicometria va tenir lloc entorn del mateix concepte de *mesura*. Inspirat per la British Association for the Advancement of Science, psicòlegs i físics van proposar-se el repte de decidir si, amb els avenços en el desenvolupament dels tests, la mesura psicològica podia ser enquadrada dintre del model general de mesura dels atributs físics. En síntesi, mesurar consistia a quantificar, és a dir, determinar la quantitat en què un atribut és present en un objecte. I per a fer-ho, la mesura dels atributs físics depenia de la capacitat d'observació de relacions entre objectes com a conseqüència d'una operació o manipulació empírica. En tant que aquesta possibilitat només po-

Per saber-ne més

L'estudiant interessat a aprofundir sobre la constitució de la psicometria com a disciplina científica des d'una perspectiva històrica pot consultar les revisions de Hambleton (1994), Brennan (1997), Traub (1997) o Bock (1997).

dia ser satisfeta amb els objectes físics, la controvèrsia va tenir una resposta inicial contundent. Tal com van concloure Ferguson i els seus col·laboradors (1940), la mesura psicològica que es proposava la psicometria simplement no era possible, ja que aquest model no es podia estendre per a incloure els atributs psicològics que, per definició, no eren observables ni manipulables empíricament. El debat, però, no va quedar tancat i va ser Stevens (1946) amb el seu treball *On the theory of scales of measurement* qui va fer un pas fonamental per al desenvolupament del que posteriorment va ser anomenat el *model representacional de mesura psicomètrica*.

Stevens va definir la mesura com el procés d'assignació de números a objectes o esdeveniments d'acord amb unes regles, producte de les quals s'obtidrien els diferents tipus d'escalles proposades: nominal, ordinal, d'interval i de raó. Mesurar, en aquest sentit, no consistiria únicament en quantificar sinó que seria el producte de la utilització de diferents regles que, en darrer terme, determinarien el tipus d'operacions –o tècniques– estadístiques permeses en cada escala. No exempta de crítiques, és important destacar la importància de la seva contribució com un primer intent per superar les restriccions imposades per la quantificació dels atributs físics amb l'objectiu de resoldre la controvèrsia entorn de la mesura dels fenòmens psicològics. Així, el debat sobre els models o paradigmes de mesura va formar part important de l'agenda psicomètrica i es van posar les bases per al desenvolupament posterior d'altres models com l'operacional i el clàssic, sobre els quals tornarem més endavant. D'altra banda, tancat aquest breu repàs de les contribucions importants durant els anys quaranta, no podem deixar de fer menció de la publicació de Thurstone (1947) de l'influent *Multiple factor analysis*, que a partir dels treballs de Spearman, Kelley i Burt dona el suport estadístic necessari per a la construcció i validació dels tests durant les dècades següents.

En la dècada dels anys cinquanta assistim a la publicació de les altres dues obres de referència per a la psicometria moderna. Instal·lada en el circuit acadèmic, la feina portada a terme per diversos autors convergeix en el que s'ha anomenat la *teoria clàssica dels tests* (TCT). I ho fa de la mà de Gulliksen (1950) en el seu *Theory of mental tests*, en què formalitza el model lineal clàssic per primera vegada i en defineix les assumpcions principals. Com a teoria dels tests, la TCT proposa un nou enfocament basat en el concepte de puntuació vertadera. Partint de la puntuació empírica obtinguda mitjançant els tests, i seguint un conjunt determinat de supòsits, l'objectiu és descompondre-la en dues parts fonamentals per tal d'estimar l'error associat al procés de mesura i, així, inferir el valor real que es vol mesurar. No sense dificultats, es va convertir en poc temps en la teoria dels tests de referència i va estimular tant el debat sobre la mateixa mesura psicològica com el procés de desenvolupament dels tests durant les dècades següents. De manera anàloga, Torgerson (1958) publica *Theory and methods of scaling* i estableix el cànon per a l'escalament psicofísic i psicològic, és a dir, l'ordenació dels estímuls de manera paral·lela a la de les persones.

Finalment, els anys seixanta són una dècada de **revisió crítica i obertura de nous camins** per a la psicometria. A partir de la TCT, el debat entorn de l'estimació dels errors va cristal·litzar en dos nous corrents en el desenvolupament de tests: la teoria de la generalitzabilitat i la teoria de resposta a l'ítem. Per tal d'incrementar la precisió de les mesures, Cronbach i els seus col·laboradors⁵ van proposar la teoria de la generalitzabilitat, que mitjançant l'aplicació de l'anàlisi de variància permetria descompondre l'error genèric proposat per la TCT en la cerca dels seus diferents elements. Així, d'acord amb aquesta teoria, l'anàlisi de la fiabilitat es basa en el disseny d'investigacions que permeten analitzar les diferents fonts d'error –facetes, segons els seus termes– que afectarien el procés de mesura. Entre aquestes, per exemple, la forma del test, els ítems que el componen, les ocasions en què s'administra o els participants. Tot i que va ser una innovació important, la complexitat d'aquesta teoria va limitar la seva difusió a la pràctica i en poc temps els seus avenços van ser bàsicament reformulats per un segon enfocament, la teoria de resposta a l'ítem (TRI).

⁽⁵⁾Cronbach, Rajaratnam i Gleser (1963); Gleser, Cronbach i Rajaratnam (1965).

Així, la TRI es presenta com la resposta a les crítiques principals que havia rebut la TCT. En aquest sentit, és convenient fer esment de la publicació per part de Lord i Novick (1968) del seu *Statistical theories of mental test scores*. En aquest treball, també amb esperit de síntesi i anàlisi crítica de la feina desenvolupada en el context de la TCT, es plantegen les dificultats que aquest model no estava essent capaç de resoldre. Bàsicament, els problemes principals podrien ser organitzats entorn de la dependència dels instruments i les mostres utilitzades en el procés de construcció i administració dels tests. Durant les dècades posteriors la TRI ocuparà, juntament amb la TCT, un lloc privilegiat entre les teories dels tests desenvolupades per la psicometria. Tornarem més endavant sobre aquestes qüestions però val la pena tancar aquesta etapa daurada de la psicometria moderna assenyalant un últim desenvolupament que també va tenir el seu inici en la dècada dels anys seixanta. Es tracta dels tests referits a criteri que, en el context de l'educació, tenen com a objectiu avaluar la destresa de les persones en un camp de coneixement molt ben delimitat.

Avaluar en funció de criteris i no sobre la base de la norma –és a dir, escalar o ordenar els individus comparant-ne les puntuacions– no és un enfocament nou. Al contrari, com hem vist en el recorregut històric del naixement i establiment de la psicometria com a disciplina científica, aquesta forma d'avaluació ha estat objecte d'interès amb anterioritat, però és gràcies a les contribucions de Glaser (1963) i Popham i Husek (1969) que el treball en aquest context queda formalitzat. El que és innovador, des del punt de vista del desenvolupament dels tests, és l'enfocament en el procediment per a establir els estàndards d'avaluació i també en la consistència i precisió amb què es classifiquen els individus d'acord amb aquests estàndards. Tot i que va néixer en el marc de la TCT, aplicacions posteriors de la TRI han permès aprofitar els avenços en el desenvolupament dels tests referits a criteri com a instruments de mesura dels fenòmens psicològics.

1.3. La psicometria en el context de l'avaluació psicològica

Una vegada recorregut el camí dels antecedents i els orígens de la psicometria moderna, estem en disposició d'abordar-ne formalment la definició. I podem fer-ho a partir de les definicions que s'han anat proposant en els textos de referència. Aquesta tasca, però, no està exempta de risc. Els autors elaboren definicions concretes que, probablement, no reflecteixen tots els matisos amb què després les contextualitzen en els seus textos. Malgrat això, podem aprofitar aquestes propostes per a valorar els diferents aspectes en què es posa un èmfasi especial. Al cap i a la fi, no hi cap tot en una definició.

Més enllà del tòpic que diu que hi ha tantes definicions com autors, podem distingir tres grans aproximacions. Per una banda, un primer grup de definicions aborda la psicometria a partir dels **instruments que utilitza**. Des d'aquest corrent es circumscriu la psicometria com la disciplina encarregada de desenvolupar els fonaments per a la construcció i administració de tests (per exemple, vegeu Martínez Arias, Hernández Lloreda i Hernández Lloreda, 2006). Aquesta definició, tot i que curta i concisa, presenta algunes dificultats. En primer lloc, la paraula *test* és polisèmica, de manera que segons el context en què sigui emprada pot esdevenir sinònim d'altres termes amb significats diferents com *prova*, *examen* o fins i tot *assaig*. En segon lloc, més enllà de la precisió semàntica que es pot sobreentendre en el context de la psicologia, no deixa de ser en certa manera restrictiva. Si és ben cert que els tests psicològics són els instruments específics que desenvolupen i administren els psicòmetres, també ho és que són el resultat aplicat d'un procés més ampli caracteritzat pel desenvolupament de teories i mètodes orientats a la mesura indirecta dels fenòmens psicològics.

Un segon corrent defineix la psicometria a partir de l'**objecte a què sotmet o aplica el seu interès**. Així, la psicometria es defineix de manera més o menys genèrica com la disciplina científica orientada a l'avaluació o mesura dels fenòmens psicològics (per exemple, vegeu Rust i Golombok, 2009). Aquesta definició s'ajusta de fet al significat etimològic de la paraula *psicometria*, que en els seus orígens grecs podem trobar en la juxtaposició de les paraules *psique* – que significa 'ànima', 'alè' o 'intel·lecte' – i *metron* –en referència al procés de mesura. En aquesta línia es pot situar la definició de Kline (1998), que aborda específicament la psicometria com la tasca de desenvolupar mesures científiques fonamentals en les àrees de la personalitat i les capacitats. En aquest context, la científicitat es converteix en sinònim d'*estandardització* i, tot i no fer-hi referència explícita, al·ludeix indirectament a la utilització de tests en el context del mètode científic. Un altre aspecte implícit important és, amb relació a l'estatus mètric de les puntuacions obtingudes com a resultat del procés de mesura psicològica, el caràcter quantitatiu que hauria de tenir la psicometria com a disciplina científica. Com ja hem avançat en l'aproximació històrica, aquest és un tema controvertit sobre el qual haurem de tornar més endavant.

En una posició intermèdia, un tercer grup d'autors situa la psicometria en la **intersecció dels dos corrents anteriors**. En aquesta línia podem assenyalar per exemple la definició de Buchanan i Finch (2005), que especifica els dominis de mesura estandarditzada mitjançant tests amb relació a les habilitats, els atributs i les característiques psicològiques. És important assenyalar que, malgrat que en alguns casos es fa referència explícita a la utilització de tests, entre aquestes definicions podem trobar també un reconeixement explícit del sentit més ample del procés de mesura. Així, Holden (2000) defineix la psicometria com la teoria i la tècnica de mesura que, en el context de la psicologia, s'encarrega dels factors que són mesurables. En aquest sentit, Muñiz (2003) ho fa definint la psicometria com el conjunt de mètodes, tècniques i teories implicades en la mesura de les variables psicològiques tenint-ne en compte l'especialització en les propietats mètriques exigibles a aquest tipus de mesura. Finalment, Jones i Thissen (2007) descriuen la psicometria com la disciplina –quantitativa, especifiquen en aquest cas– encarregada de desenvolupar models i mètodes orientats principalment al resum, la descripció i l'establiment d'inferències a partir de les dades recollides en la investigació psicològica.

Com hem pogut observar, la definició de la psicometria s'ha desenvolupat fonamentalment entorn dels mitjans o instruments que utilitza –els tests– i l'objecte que persegueix amb el seu ús –la mesura dels fenòmens psicològics. En canvi, una aproximació comprensiva a aquesta definició hauria de considerar un tercer element important vinculat amb la **finalitat última a la qual serveix**. Tal com hem esmentat en aquest apartat, la psicometria adquireix tot el seu significat amb relació a l'àrea de la psicologia en què té el seu paper fonamental: l'avaluació psicològica. D'aquesta manera, hem de tenir present que la mesura dels fenòmens psicològics mitjançant el desenvolupament i l'administració de tests no és més que una part del procés general d'avaluació, i no sempre és la més important⁶.

⁶Podeu veure Fernández-Ballesteros (1997) per a una discussió més àmplia.

La psicometria, en aquest sentit, contribueix al desenvolupament de l'avaluació psicològica proporcionant teories, mètodes i tècniques que, en darrer terme, permeten descriure, classificar, diagnosticar, explicar o predir els fenòmens psicològics objecte de mesura. De fet, contextualitzada així la psicometria podem anar un pas més enllà i assenyalar que, al seu torn, l'avaluació psicològica no es troba en cap altre lloc que al servei de la intervenció psicològica. És a dir, la finalitat última que guia el desenvolupament i l'administració de tests és la de contribuir a la recollida de les evidències necessàries que permetin als psicòlegs prendre una decisió o orientar alguna acció. Evidentment aquesta intervenció dependrà del context en què l'avaluació hagi estat desenvolupada, però és important tenir-ho present per a entendre no solament què és la psicometria, sinó la seva importància per a l'exercici professional dels psicòlegs en els diferents contextos en què intervenen.

Partint d'aquest enfocament comprensiu de la psicometria, podem sintetitzar les diferents aproximacions i formular finalment una definició que incorpori el seu paper en el marc de l'avaluació i la intervenció psicològiques.

Així, la psicometria és una branca de la psicologia que, mitjançant teories, mètodes i tècniques vinculats al desenvolupament i l'administració de tests, s'ocupa de la mesura indirecta dels fenòmens psicològics amb l'objectiu de fer descripcions, classificacions o diagnòstics, donar explicacions o fer prediccions que permetin orientar una acció o prendre decisions sobre el comportament de les persones en l'exercici professional de la psicologia.

Per concloure aquest exercici de definició formal és important discutir breument les diferències que hi ha entre la psicometria i una altra disciplina també interessada per la mesura psicològica com és la **psicologia matemàtica**⁷. En aquest sentit, podem reconèixer la psicologia matemàtica com la branca de la psicologia interessada pel desenvolupament de models dels processos perceptius, cognitius i motors amb l'objectiu d'establir lleis que relacionin estímuls físics amb els comportaments a partir d'estudis quasi exclusivament experimentals. Són molts els autors comuns a les dues orientacions des dels orígens de l'estudi sistemàtic de les diferències humanes al segle XIX però, a diferència de la psicometria, la psicologia matemàtica no està tan interessada per les diferències individuals com per la definició de lleis generals que modelitzin el comportament mitjà de les persones.

⁽⁷⁾Podeu veure, entre d'altres, Jáñez (1989), i Padilla, Merino, Rodríguez-Miñón i Moreno (1996).

2. Fonaments de la psicometria

Abordada la psicometria des d'una perspectiva històrica, tractant les seves arrels i la seva construcció com a disciplina científica en el context de la psicologia, en aquest segon apartat en revisarem els fonaments. Començarem discutint una definició i classificació dels tests, progressarem pels diferents models de mesura psicomètrica desenvolupats en les diferents teories dels tests, i farem una breu introducció a la teoria més estesa en la pràctica actual: la teoria clàssica dels tests. Finalment, recapitularem aquests fonaments atenent al procés que la psicometria segueix per establir les seves inferències sobre els fenòmens psicològics no observables a partir de les puntuacions obtingudes mitjançant tests. Comencem, doncs, amb una definició i classificació dels tests.

2.1. Definició i classificació dels tests

D'acord amb la definició del manual clàssic d'Anastasi (1988), un test psicològic⁸ és un procediment de mesura objectiva i estandarditzada d'una mostra de comportaments. Altres definicions són possibles⁹ però aquesta conté tres elements fonamentals que ens permeten abordar sistemàticament les característiques més importants que compleixen de manera genèrica els tests. D'una manera o altra, ja les hem anat introduint en aquest text.

⁽⁸⁾En endavant, en direm un test.

⁽⁹⁾Podeu veure, entre d'altres, Kaplan i Saccuzzo (2001), Murphy i Davidshofer (2005) o Urbina (2004).

En primer lloc, la mesura mitjançant el desenvolupament i l'administració de tests és, o al menys pretén ser, **objectiva**. En aquest sentit, l'objectivitat fa referència a la substitució del judici personal basat en criteris subjectius per un conjunt de normes determinades i conegudes que permeten obtenir i interpretar les puntuacions dels individus en igualtat de condicions. Així mateix, la mesura que cerquen els tests és **estandarditzada**, en tant que les puntuacions obtingudes depenen d'un procediment establert d'administració, correcció i interpretació que les fan, o les haurien de fer, invariants del professional que administra els tests, les condicions específiques en què ho fa, i la manera en què obté i interpreta les puntuacions resultants de la mesura. Finalment, els tests adrecen els fenòmens psicològics no observables mitjançant una **mostra de comportaments**. En la mesura que aquesta mostra sigui representativa del conjunt, les puntuacions obtingudes permetran als professionals establir adequadament les seves inferències sobre el comportament general de les persones més enllà dels elements específics avaluats mitjançant tests.

Ja hem discutit la naturalesa no observable dels fenòmens psicològics i la dificultat intrínseca que representa per a la **mesura indirecta mitjançant tests** que es proposa la psicometria. Però és important destacar aquesta qüestió quan la comparem amb altres tipus de mesura científica. En aquest sentit, la mesura dels fenòmens psicològics no resulta tan senzilla com la de les magnituds físi-

ques dels objectes –per exemple, el seu pes o la seva longitud–, que les ciències naturals poden observar i manipular directament. A més a més, tal com assenyala García Cueto (1993), hi ha d'altres **diferències importants** entre aquest tipus de mesura i la mesura indirecta dels fenòmens psicològics que han de ser tinguts en compte.

D'una banda, l'objectiu habitual de la mesura de les magnituds físiques és obtenir informació sobre un únic objecte. La psicometria, en canvi, és proposa desenvolupar instruments que permetin obtenir-la sobre un grup d'individus amb l'objectiu de treure conclusions sobre cadascun d'ells, sobre el grup sencer, i fins tot extrapolar-ne els resultats a les poblacions de referència d'on provenen aquests individus. D'altra banda, la mesura de magnituds físiques parteix de la possibilitat de repetir-ne el procediment tantes vegades com es vulgui sense variar les condicions en què es porta a terme la mesura. Aquest procediment és incompatible amb la mesura mitjançant tests, en tant que la repetició en l'aplicació d'una mateixa prova sobre els mateixos individus produeix variacions en les puntuacions que poden ser explicades pel cansament o per l'efecte de la pràctica en l'aprenentatge i no per variacions substancials en els fenòmens d'interès. Malgrat això, podem trobar algunes pràctiques desenvolupades en els contextos aplicats de les ciències naturals que tenen molt en comú amb l'administració de tests.

Així, quan es vol analitzar la contaminació atmosfèrica d'una ciutat o el nivell d'alcohol en sang d'una persona, els científics desenvolupen un conjunt de proves objectives –anomenades *reactius*– per tal d'aplicar-les de manera estandaritzada sobre una mostra representativa de l'objecte que volen mesurar –tant si és l'aire de la ciutat com la sang de l'individu, seguint els mateixos exemples. En aquest sentit, el procés de desenvolupament i administració de tests és comparable amb aquestes pràctiques, i posa en relleu el sentit que la psicometria atribueix als tests com els seus instruments de mesura. Un test, lluny del domini de la psicologia, no és més que una prova, un examen o un assaig. En definitiva, és un **reactiu** que, en el cas de la mesura indirecta dels fenòmens psicològics mitjançant tests, no té cap altra finalitat que produir una reacció en el comportament dels individus per a registrar-lo i obtenir una puntuació com a resultat del procés de mesura. Aquest sentit etimològic de la paraula *test* el podem trobar en l'arrel de la seva definició en el context de la psicometria (per exemple, vegeu Yela, 1984), en què s'assimila directament amb el reactiu de les ciències naturals i se n'espera que, aplicat a un individu, reveli i doni testimoni fidel dels fenòmens psicològics no observables que són objecte de mesura.

Des d'aquest punt de vista, el valor d'un test es troba en la capacitat per a **suscitar i mesurar comportaments** que resultin un bon indicador –és a dir, una bona representació– del conjunt global de comportaments implicats en els fenòmens d'interès. En aquest sentit, hem de fer una precisió important amb relació als fenòmens que són objecte de mesura mitjançant tests. En aquest text parlem dels fenòmens psicològics d'una manera general i ho fem en tant

que l'objectiu d'un test és la mesura objectiva i estandarditzada d'una mostra de comportaments. Això no vol dir, però, que els tests hagin d'adreçar-se exclusivament a elements amb una llarga tradició en la teoria psicològica com són, entre d'altres, la intel·ligència o la personalitat. Per exemple, alguns tests s'adrecen a la mesura de les respostes fisiològiques, com poden ser la conductivitat de la pell o la freqüència cardíaca. D'altres es fixen en qüestions com les actituds racistes, el consum de substàncies o la sociabilitat de les persones. Finalment, en el camp de l'educació, no són pocs els tests que tenen per objectiu avaluar el domini dels estudiants en diferents àrees de coneixement. En tots el casos parlem de fenòmens psicològics en tant que responen a objectes d'interès per a les ciències socials i del comportament i, a més a més, recauen en els dominis de la mesura indirecta mitjançant tests sempre que es reunixin les condicions que estableix la psicometria.

Així, la mesura indirecta que es proposen els tests comença amb una definició precisa dels fenòmens objecte d'interès i se sustenta fonamentalment en dos tipus de teories. D'una banda, en una **teoria substantiva** sòlida i ben establerta sobre el comportament de les persones –per exemple, una teoria de la intel·ligència–, que donarà el suport teòric necessari per a definir els elements crítics que conformen aquests fenòmens i determinarà els comportaments implicats que seran emprats com a evidències observables en el procés de mesura. De l'altra, en una **teoria dels test**, que permet establir les inferències sobre els fenòmens psicològics no observables a partir de les puntuacions obtingudes en el procés de mesura. Al seu cor, un **model de mesura** determinat que serveix al propòsit de relacionar les puntuacions obtingudes i els fenòmens objecte de mesura, a partir del qual s'articula el **procés d'inferència psicomètrica**. Tant els models de mesura com el mateix procés d'inferència psicomètrica seran tractats en detall més endavant. Però és important fer-ne esment en aquest punt perquè resulten clau per a entendre la mesura indirecta que es proposen els tests i que els defineix amb relació a d'altres tipus de mesura científica.

Finalment, una última consideració sobre els tests i els diferents termes amb què es fa referència amb ells ens porta a fixar-nos en dues grans formes de tractar les respostes obtingudes per mesurar els fenòmens psicològics d'interès. D'una banda, els anomenats **tests d'habilitat o de potència** tenen com a objectiu avaluar la competència, l'aptitud o els coneixements dels individus a partir de l'encert o la qualitat de les seves respostes. Són proves que discriminen respostes correctes i incorrectes i és aquesta la base per a puntuar les execucions individuals. De l'altra, els anomenats **tests de personalitat**¹⁰ tenen un objectiu diferent i cerquen conèixer de manera general les motivacions, preferències, opinions o actituds dels individus envers a un determinat estímul. Aquest segon tipus de tests no tenen respostes correctes i per tant no serveixen per a avaluar l'encert o l'error dels individus. Aquesta distinció és rellevant des d'un punt de vista terminològic, ja que els tests de personalitat són moltes vegades anomenats **qüestionaris, inventaris o escales**, malgrat que aquest ús no és sempre consistent. De fet, el terme *escala* pot esdevenir sinònim de *test* quan aquest està conformat per diferents parts, o fins i tot de cadascuna d'aquestes

⁽¹⁰⁾Tot i que el terme personalitat pot fer referència a la branca de la psicologia interessada per les característiques personals que influeixen les seves cognicions, emocions, motivacions i comportaments, en aquest cas s'utilitza en un sentit més ampli per a fer referència als tests que, en oposició als d'habilitat o de potència, no pretenen mesurar les capacitats de les persones.

parts per a reflectir les dimensions o característiques específiques que mesuren en el context del test global. D'altra banda, un altre terme estès també és el de **bateria**, que tampoc no té un significat unívoc i pot fer referència tant a un test compost per diverses parts com a una selecció de diferents tests administrats conjuntament per un professional en una avaluació psicològica determinada.

Després d'aquesta discussió terminològica, i malgrat l'enorme diversitat de tests que hi ha actualment, estem en disposició de classificar-los en funció d'algunes de les seves característiques més importants. Així:

- **Segons el propòsit.** El tests poden tenir diferents finalitats, i entre elles en podem destacar dues de fonamentals: la **diagnosi**, orientada a l'avaluació de les condicions actuals dels individus, i la **predicció**, que es proposa relacionar la mesura actual amb el comportament de les persones en situacions futures. En aquets sentit, per exemple, una cosa és obtenir informació sobre la destresa d'una persona en la resolució de problemes i una altra és utilitzar aquesta informació per a predir el seu desenvolupament en una feina determinada. Altres propòsits més específics són també possibles, com ja hem comentat en la discussió de la definició formal de la psicometria.
- **Segons el contingut.** De manera general, podem classificar els tests segons l'àrea de la psicologia a què pertanyen els fenòmens psicològics que volem mesurar. Aquesta classificació, però, varia segons els autors que la duen a terme, en tant que es poden fixar en diferents nivells de complexitat en jerarquitzar aquestes àrees. En el nivell més general podem distingir bàsicament tres grans grups que inclourien els tests orientats a l'avaluació de les **habilitats cognitives, la personalitat i les actituds**.
- **Segons el format.** Tenint en compte els materials utilitzats és freqüent distingir els tests de llapis i paper, de manipulació i de mesures fisiològiques. Ens referim als **tests de llapis i paper** quan presenten els ítems o preguntes en paper i requereixen de l'individu que doni algun tipus de resposta escrita. Els **tests computeritzats**, malgrat que no fan servir el paper, podrien ser encabits en aquesta categoria en tant que exigeixen algun tipus de resposta escrita mitjançant dispositius electrònics. D'altra banda, són **tests de manipulació** els que presenten una sèrie d'objectes, imatges o trencaclosques que els individus han de resoldre per a demostrar la seva habilitat. Finalment, els **tests de mesures fisiològiques** utilitzen sensors de diversa mena per a registrar les reaccions dels individus davant dels estímuls físics presentats.
- **Segons el tipus d'administració.** Una altra manera de classificar els tests pot tenir en compte la manera en què són utilitzats i ens permet distingir els **tests individuals i grupals**. Aquesta classificació se superposa a la que hem fet segons el format però ens permet distingir tests que requereixen una administració a un únic individu o permeten utilitzar-lo amb un grup

d'individus al mateix temps. També podem distingir els **tests verbals i no verbals**, en funció de si es fa una presentació oral o escrita del test i les seves instruccions o no.

- **Segons el tractament de les respostes.** Seguint la diferència que establíem entre els **tests d'habilitat o de potència** i els **tests de personalitat**, podem classificar també els tests en dos grans grups en funció de si avaluen l'encert en les respostes per a determinar la competència, l'aptitud o els coneixements dels individus, o si tracten d'avaluar motivacions, preferències, opinions o actituds. Aquests dos tipus de proves són també conegudes com a **proves d'execució màxima** i **proves d'execució típica**, respectivament.
- **Segons la interpretació de les puntuacions.** Tal com ja hem comentat, les puntuacions obtingudes mitjançant tests poden ser interpretades sobre la base de la norma o en referència a un criteri. Així, els **tests normatius** permeten comparar la puntuació de l'individu amb l'execució observada en un grup de referència que prèviament ha respost al mateix test. Aquest grup pot estar compost mitjançant mostreig probabilístic, quan representa la població de referència, o no probabilístic, quan estem interessats a comparar les puntuacions amb un grup de persones que compleix unes característiques determinades. D'altra banda, els **tests referits a criteri** prenen com a referència la definició d'un domini de coneixements o habilitats específics i permeten mesurar l'execució de l'individu, no ja en comparació amb un grup de referència, sinó en funció del seu grau d'adequació o consecució d'aquest criteri.
- **Segons l'estatus comercial.** Finalment, els tests es poden classificar en funció de si són propietaris o no. Els **tests propietaris o comercials** són tests que requereixen el pagament pel seu ús i inclouen els exemplars del mateix test i un manual que conté informació sobre el seu desenvolupament i les seves propietats psicomètriques, els fulls de correcció de les respostes i les taules per a interpretar les puntuacions amb relació als grups de referència. En d'altres casos parlem de **tests oberts o no comercials**, i generalment es poden obtenir contactant amb l'autor, el qual, habitualment, també ha desenvolupat alguna mena de manual que contextualitza el test. De vegades aquesta informació queda reduïda a la publicació d'un article científic en què l'autor l'introdueix, fa una primera administració i presenta les seves propietats psicomètriques.

2.2. Models de mesura psicomètrica

Qualsevol tipus de mesura científica, i la mesura indirecta dels fenòmens psicològics mitjançant tests no n'és una excepció, es fonamenta en una **definició del mateix concepte de mesura**. En aquest sentit, tal com ja hem introduït, en el nucli de les teories dels tests podem trobar un conjunt de models que

estableixen què és exactament mesurar i, com a conseqüència, especifiquen la relació entre les puntuacions empíriques obtingudes i els fenòmens psicològics no observables objecte d'interès. Són els **models o paradigmes de mesura**, que la psicometria ha desenvolupat de manera formal a partir del debat iniciat als anys quaranta sobre la viabilitat de la mesura dels fenòmens psicològics. Abans de caracteritzar aquests models, però, reprendrem la controvèrsia a partir de les conclusions de Ferguson i els seus col·laboradors (1940) per a la British Association for the Advancement of Science. Com veurem a continuació, representen un punt de partida important des del qual la psicometria ha desenvolupat els seus propis models de mesura per abordar específicament els fenòmens psicològics no observables.

D'acord amb el treball de Campbell (1920), físic i membre de la comissió que va elaborar aquest informe, el **model general de mesura desenvolupat en la física** es basaria en l'existència d'una equivalència –un isomorfisme, en les seves paraules– entre alguns atributs físics i les propietats additives dels números. Mesurar seria assignar números a aquests atributs per a representar l'additivitat dels objectes físics i aquesta representació seria només possible en tant que es satisfessin dues condicions: que l'atribut es pogués ordenar, és a dir, que fos possible determinar que l'atribut present en un objecte és menor, igual o més gran que en un altre; i que es pogués demostrar empíricament la presència de l'additivitat mitjançant la manipulació empírica. És el que va anomenar *mesura fonamental*, definint la mesura dels atributs físics per analogia amb els números a partir de les operacions portades a terme sobre els objectes per a quantificar-ne els atributs.

En un exemple senzill, el pes de dos o més objectes pot ser observat empíricament mitjançant una balança per a ordenar-los de més petit a més gran, com també es pot demostrar empíricament que el pes obtingut després de posar aquests objectes junts a la balança dona com a resultat la suma dels pesos parcials.

Campbell (1928) va estendre aquest model general al que va anomenar *les mesures derivades*, en què la quantificació de certs atributs físics no dependria de la manipulació empírica sinó del descobriment de relacions matemàtiques entre dues mesures fonamentals. Per exemple, seria el cas de la densitat dels objectes, mesura derivada que respondria a una raó entre la seva massa i el seu volum. En qualsevol cas, la clau de la mesura per al model de Campbell es trobaria en la necessitat d'observar i manipular empíricament els objectes, bé per a quantificar un atribut físic mitjançant la seva observació directa o bé per a establir funcions matemàtiques que relacionin altres atributs observables. En la mesura en què els fenòmens psicològics no poden ser observats ni manipulats empíricament, ni són producte d'una relació entre d'altres atributs observables, la mesura indirecta que persegueix la psicometria no seria equiparable a la mesura fonamental o derivada de la física i, per tant, resultaria inviable. Aquesta va ser la conclusió final de la comissió organitzada per la British Association for the Advancement of Science, que, com ja hem comentat, lluny de tancar la controvèrsia va obrir un dels debats més importants per a la mesura indirecta dels fenòmens psicològics mitjançant tests.

Com a resposta a les restriccions imposades pel model general de mesura desenvolupat en la física, la psicometria va avançar en el debat sobre l'aproximació indirecta als fenòmens psicològics desenvolupant les seves pròpies teories i models de mesura. Tot i que altres autors utilitzen altres termes¹¹ en aquest text seguirem a Michell (1986) i els anomenarem *model representacional*, *operacional* i *clàssic*. A continuació, caracteritzarem aquests models i analitzarem breument les diferències en el seu plantejament del procés de mesura per tal de fer-lo extensiu als fenòmens psicològics. Essent un debat encara obert, i susceptible de diferents interpretacions segons els diferents autors, optem per l'exposició de Michell en tant que planteja, des d'una visió crítica, les incògnites més importants que la psicometria ha de resoldre en la seva construcció com a disciplina científica.

En primer lloc, el **model representacional** va ser desenvolupat a partir dels treballs de Stevens (1946) i Suppes (1951), i defineix la mesura com el procés d'assignació de números a objectes a partir d'unes regles de manera que reflectixin relacions empíriques entre els objectes. Aquestes relacions, de manera comparable al model de Campbell, quedarien representades per les propietats dels números però no se centrarien exclusivament en l'additivitat. De fet, Stevens va obrir la porta a altres tipus de relacions per a portar a terme la mesura, acceptant per exemple l'equivalència o l'ordre. Mesurar no seria únicament quantificar sinó la representació numèrica dels fets empírics en sentit ample, i donaria com a resultat quatre escales clàssiques per a la psicologia segons el tipus de relació representada: escala nominal, ordinal, d'interval i de raó.

Tal com fa el model desenvolupat en la física, el model representacional assumeix la correspondència entre les propietats dels números –les puntuacions obtingudes mitjançant tests, en aquest cas– i les relacions dels objectes que representen. En canvi, trasllada el focus des de la demostració empírica d'aquestes relacions a l'establiment d'un conjunt de restriccions sobre el tipus d'operacions estadístiques permeses a cada escala que garanteixi la invariància de les relacions empíriques representades. És el debat sobre l'estadística permesa, que ha portat a una gran controvèrsia entre els partidaris del model representacional i l'operacional.

En segon lloc, el **model operacional** es basa en les contribucions de Bridgman (1927) i Dingle (1950), a partir de les quals es va proposar la definició de qualsevol concepte mitjançant les operacions necessàries per a mesurar-lo. Així, d'acord amb aquest model, la mesura no seria una altra cosa que el conjunt d'operacions necessàries per a definir un concepte que, en darrer terme, acaben produint números. Aquesta definició de la mesura té punts en comú amb la del model representacional, ja que assignar números –les puntuacions obtingudes mitjançant tests– d'acord amb unes regles és un tipus particular de les operacions possibles, i els números resultants no són altra cosa que el pro-

⁽¹¹⁾Per exemple, podeu veure Fraser (1980), Swistak (1990) o Hand (1996).

Per saber-ne més

L'estudiant interessat pot trobar una discussió molt més àmplia de la que podem plantejar en aquest text a Michell (1999).

Per saber-ne més

En relació amb el debat sobre l'estadística permesa, podeu veure, entre d'altres, la interessant discussió de Gaito (1980), i Velleman i Wilkinson (1993).

ducte de les operacions portades a terme. En canvi, una diferència important entre aquests dos models de mesura es troba en el requeriment segons el qual els números representen o no un sistema de relacions empíriques.

Per al model representacional, aquestes relacions empíriques són prèvies a la mesura i tenen una existència independent de les operacions portades a terme per a produir els números. El model operacional, en canvi, limita el domini de les operacions a les que d'una manera consistent produeixen números i, en cap cas, està interessat per l'existència d'una realitat empírica que doni suport a aquests números. Aquesta diferència en l'enfocament de la mesura, a més, té implicacions importants quant al debat sobre l'estadística permesa, ja que els partidaris del model operacional no veuen la necessitat d'establir cap restricció sobre les operacions estadístiques. En no ser l'objectiu la cerca de relacions empíriques comparables a les relacions entre els números, sinó els números en sí mateixos, cap restricció no pot ser aplicable a unes puntuacions que, en paraules de Lord (1953), no tenen consciència d'on provenen.

Finalment, el **model clàssic** es proposaria resoldre aquesta discussió plantejant la naturalesa quantitativa dels fenòmens psicològics com a condició per a la mesura. Partint dels treballs de Rozeboom (1966) i Jones (1971), i arrelat en el desenvolupament de les teories quantitatives de la psicologia de les dècades anteriors, aquest model defineix la mesura com la determinació de la quantitat –del quant– en què un atribut és present en l'objecte mesurat. És a dir, mesurar consisteix a determinar quantes unitats són presents en l'atribut observat però, a diferència dels models de Campbell i Stevens, no exigeix que hi hagi una relació empírica entre els objectes. De fet, el model clàssic mesura atributs, no objectes, i per tant no preveu en el seu nucli la demostració de l'additivitat mitjançant l'observació i la manipulació empírica dels objectes físics. Encara més, a diferència dels models representacional i operacional, la mesura no es considera l'assignació de números a objectes d'acord amb unes regles ni el resultat d'un tipus particular d'operacions que produeix números, respectivament, sinó que es defineix com el procés de descobriment de relacions numèriques entre els valors observats en un atribut quantitatiu.

Seguint l'exemple del pes de dos o més objectes físics, les seves mesures en una unitat determinada (per exemple, en quilos) proporcionen un sistema per a relacionar-los matemàticament en termes de les seves magnituds relatives, i permeten afirmar que un d'ells pesa n vegades més o menys que un altre. Aquestes relacions numèriques són, per al model clàssic, tan vàlides com ho són les relacions observables entre els objectes i per tant comparteixen el mateix estatus com a evidències empíriques necessàries per al desenvolupament científic. En aquest sentit, és molt important assenyalar el paper de la teoria substantiva, ja que és la base sobre la qual es fonamenten tant la hipòtesi sobre la naturalesa quantitativa dels fenòmens psicològics objecte de mesura com el procés de determinació de les relacions numèriques amb les mesures obtingudes. Finalment, el model clàssic no preveu les diferents escales de mesura acceptades pels models representacional i operacional. D'aquesta manera les

escales nominals i ordinals queden fora del domini de la mesura –que, recordem, és quantitativa per definició– i per tant no és necessària cap prescripció entorn del debat sobre l'estadística permesa.

Com ja hem assenyalat, els models de mesura psicomètrica es troben a la base de les diferents teories dels tests i durant les darreres dècades han proporcionat diverses alternatives per a l'aproximació indirecta a la mesura dels fenòmens psicològics que es proposa la psicometria. Així, aquests tres models de mesura –dèiem, representacional, operacional i clàssic– es corresponen amb la **teoria representacional**, la **teoria clàssica dels tests** i els **models de variable latent**, respectivament. Abordar totes aquestes teories alternatives està fora dels objectius d'un text introductor com aquest, tot i que l'estudiant interessat pot trobar una anàlisi molt detallada per exemple a Borsboom (2005). En canvi, optarem per restringir el tractament de les teories dels tests a la que gaudeix del suport majoritari en la pràctica actual de la psicometria. Parlem de la teoria clàssica dels tests (TCT), que és, a més a més, la base per al desenvolupament dels mòduls posteriors d'aquest text. És important destacar, per a evitar confusions, que la TCT no s'anomena així pel model de mesura psicomètrica que utilitza. De fet, la seva definició de mesura encaixa amb el model operacional i, en canvi, adopta aquest nom per oposició a d'altres enfocaments més moderns com la teoria de resposta a l'ítem (TRI). Finalment, introduïrem molt breument algunes de les innovacions que, d'acord amb el model clàssic de mesura propi dels models de variable latent, la TRI ha desenvolupat per tractar de resoldre algunes de les limitacions de la TCT.

2.3. Teoria clàssica dels tests

La TCT és la teoria dels tests més estesa actualment en la pràctica de la psicometria i es basa en el **model lineal clàssic** proposat per Spearman¹², sistematitzat per Gulliksen (1950) i reformulat posteriorment per Lord i Novick (1968), que articula el procés de mesura definint tres conceptes fonamentals: la puntuació vertadera, la puntuació empírica, i l'error de mesura. Per a introduir-lo, en aquest text seguirem l'exposició de Muñiz (1996 i 2003).

Així, partint del model operacional de mesura psicomètrica, aquesta teoria dels tests no està interessada pel sistema de relacions empíriques sinó que centra la seva atenció en l'anàlisi de les puntuacions obtingudes per tal d'estimar els errors comesos en el procés de mesura indirecta dels fenòmens psicològics. És l'anomenada **puntuació empírica (X)**, que, d'acord amb aquesta teoria, respondria a una relació lineal de dos components fonamentals:

$$X = V + e$$

⁽¹²⁾Per a una discussió històrica, podeu veure Traub (1997).

Per saber-ne més

L'estudiant interessat pot trobar explicacions més detallades a Borsboom (2005), Crocker i Algina (2006) o De Gruijter i Van der Kamp (2008).

D'una banda, la **puntuació vertadera** (V), que seria el resultat ideal o desitjat en què el procés de mesura mitjançant tests hauria estat portat a terme lliure de qualsevol tipus d'error. De l'altra, l'**error de mesura** (e), que seria responsable de la discrepància entre la puntuació vertadera que es vol aconseguir i la puntuació empírica obtinguda com a resultat de l'administració del test.

Com en qualsevol tipus de mesura científica, la mesura indirecta dels fenòmens psicològics es troba subjecte a aquestes variacions o desviacions no desitjades, i el model lineal clàssic proposa un conjunt de supòsits, que permeten fer una estimació de les puntuacions vertaderes a partir de les puntuacions empíriques obtingudes, i una definició de tests paral·lels. D'aquesta manera:

- **Primer supòsit:** $V = E(X)$. La puntuació vertadera (V) es defineix matemàticament com l'esperança matemàtica de la puntuació empírica (X). És a dir, el primer supòsit del model lineal clàssic assumeix que, en cas que fos possible l'administració d'un test un nombre infinit de vegades, la mitjana de les puntuacions empíriques obtingudes ens donaria com a resultat la puntuació vertadera del subjecte en el test. La puntuació empírica, per tant, no és un substitut de la puntuació vertadera sinó la millor aproximació disponible en un procés de mesura que, de manera explícita, és reconegut com a no lliure d'error.
- **Segon supòsit:** $\rho(v, e) = 0$. No hi ha correlació entre les puntuacions vertaderes dels subjectes (v) en un test i els errors respectius de mesura (e). És a dir, d'acord amb el segon supòsit, no s'espera que la grandària dels errors comesos estigui sistemàticament associada a la grandària de les puntuacions vertaderes.
- **Tercer supòsit:** $\rho(e_j, e_k) = 0$. Si disposem de dos tests diferents (j i k), no hi ha correlació entre els errors de mesura comesos amb cadascun d'ells (e_j i e_k , respectivament). És a dir, d'acord amb el tercer supòsit, els errors de mesura dels diferents tests són aleatoris cada vegada i, per tant, no s'espera que hi hagi cap relació entre ells.
- **Definició de tests paral·lels:** $V_j = V_k$ i $\sigma^2(e_j) = \sigma^2(e_k)$. Finalment, dos tests (j i k) són paral·lels sempre que les seves puntuacions vertaderes (V_j i V_k) i les seves variàncies dels errors de mesura ($\sigma^2(e_j)$ i $\sigma^2(e_k)$, respectivament), siguin idèntiques.

D'aquest model, amb els seus supòsits i la definició de tests paral·lels, es deriven un conjunt de deduccions immediates importants que formen la base per al mòdul dedicat a la fiabilitat en aquest text. Així:

- $e = X - V$. D'acord amb la formulació inicial del model lineal clàssic, l'error de mesura seria la diferència entre la puntuació empírica i la puntuació vertadera.
- $E(e) = 0$. L'esperança matemàtica dels errors de mesura és 0, de manera que, si fos possible administrar un test un nombre infinit de vegades, aquests errors aleatoris o no esbiaixats es compensarien o anul·larien entre ells.
- $\mu_x = \mu_v$. La mitjana de les puntuacions empíriques (μ_x) és igual a la mitjana de les puntuacions vertaderes (μ_v).
- $\text{cov}(V, e) = 0$. D'acord amb el segon supòsit del model, les puntuacions vertaderes no covarien amb els errors de mesura.
- $\text{cov}(X, V) = \text{var}(V)$. La covariància entre les puntuacions empíriques i vertaderes és igual a la variància de les puntuacions vertaderes.
- $\text{cov}(X_j, X_k) = \text{cov}(V_j, V_k)$. La covariància entre les puntuacions empíriques de dos tests (X_j i X_k) és igual a la covariància de les seves puntuacions vertaderes (V_j i V_k).
- $\text{var}(X) = \text{var}(V) + \text{var}(e)$. La variància de les puntuacions empíriques és el producte de la suma de la variància de les puntuacions vertaderes i la dels errors de mesura.
- $\rho(X, e) = \sigma_e / \sigma_x$. La correlació entre les puntuacions empíriques i els errors de mesura és el resultat de la divisió de la desviació típica dels errors entre la de les puntuacions empíriques.
- $\mu_1 = \mu_2 = \dots = \mu_k$. Per a k tests paral·lels, les mitjanes són idèntiques.
- $\sigma^2(X_1) = \sigma^2(X_2) = \dots = \sigma^2(X_k)$. De la mateixa manera, les variàncies de k tests paral·lels són idèntiques.
- $\rho(X_1, X_2) = \rho(X_1, X_3) = \dots = \rho(X_j, X_k)$. Finalment, les correlacions entre k tests paral·lels són també idèntiques.

Tenint en compte que ni les puntuacions vertaderes ni els errors de mesura són observables directament a partir de les puntuacions empíriques, cap dels supòsits ni les vuit primeres deduccions immediates són demostrables.

Com ja hem discutit anteriorment, la TCT resulta una aproximació útil per a l'estimació de les puntuacions vertaderes a partir de les puntuacions empíriques, però no està exempta d'**algunes limitacions** que, des del model lineal clàssic, no són fàcils de resoldre. Tal com van discutir Lord i Novick (1968) en la seva reformulació, aquesta teoria té dues dependències importants amb re-

Per saber-ne més

L'estudiant interessat pot trobar el desenvolupament matemàtic que sustenta aquestes deduccions a Muñiz (2003).

lació als instruments i les mostres utilitzades. D'una banda, una dependència de les puntuacions empíriques obtingudes respecte dels instruments, de manera que, per exemple, dos tests d'intel·ligència independents utilitzen escales diferents i, per tant, les puntuacions que se n'obtenen no resulten comparables directament. De l'altra, una dependència de les propietats psicomètriques dels tests respecte de les mostres utilitzades per a desenvolupar-los, de manera que, per exemple, la dificultat dels ítems que conformen un test depèn de les característiques dels individus a qui s'administra el test.

Tractant de resoldre aquests problemes, la TRI es proposa canviar el focus des del tractament del test sencer al tractament individual dels ítems. Per fer-ho, aquesta teoria dels tests planteja l'existència d'una relació entre les puntuacions en la variable latent –recordem, d'acord amb el model clàssic de mesura psicomètrica– i la probabilitat d'encertar cada ítem introduint diferents funcions matemàtiques per a modelitzar adequadament aquesta relació. Així, una vegada seleccionada la funció més adequada es construeixen el que s'anomenen *les corbes característiques dels ítems*, una modelització de les respostes que pot tenir en compte diferents paràmetres com la capacitat de discriminació dels ítems, la seva dificultat o la probabilitat que siguin encertats a l'atzar.

No és aquest el lloc per a fer una discussió en profunditat, però sí per a fer èmfasi en les implicacions que tenen els diferents models de mesura psicomètrica a partir dels quals la TCT i la TRI es proposen la mesura indirecta mitjançant la utilització de tests. D'una banda, tal com fa la TCT, definint les puntuacions empíriques com la suma de les puntuacions vertaderes i l'error de mesura i, de l'altra, tal com fa la TRI, establint diferents funcions matemàtiques segons el model de relació entre la variable latent i la capacitat de discriminació dels ítems.

2.4. El procés d'inferència psicomètrica

Per concloure amb els seus fonaments, recapitularem el procés que la psicometria segueix per establir-ne les inferències sobre els fenòmens psicològics no observables a partir de les puntuacions obtingudes mitjançant tests. Abans d'abordar-ne les especificitats és important tenir present que la mesura objectiva i estandarditzada d'una mostra de comportaments s'ajusta, de manera general, al procediment establert pel **mètode científic**. El podríem resumir així:

- 1) Formular una pregunta d'investigació o una hipòtesi rellevants.
- 2) Especificar i definir adequadament totes les variables involucrades.
- 3) Desenvolupar o escollir els instruments i procediments necessaris per a portar a terme les mesures.

Per saber-ne més

L'estudiant interessat pot trobar una comparació molt entenedora d'aquestes dues teories dels tests a Muñiz (2010), com també desenvolupaments més detallats de la TRI a Muñiz i Hambleton (1997), Maydeu-Olivares i McArdle (2005), Borsboom (2005), o Martínez Arias, Hernández Lloreda i Hernández Lloreda (2006).

Una referència clàssica sobre els fonaments de la TRI és la de Hambleton, Swaminathan i Rogers (1991).

4) Avaluar el funcionament dels instruments i procediments per a obtenir les garanties suficients sobre la qualitat del procés de mesura.

5) Recollir les evidències necessàries que permetin respondre als objectius de la investigació.

6) Resumir i, sempre que sigui possible, tractar estadísticament les dades obtingudes per a determinar fins a quin punt els resultats són significatius i, per tant, no són producte de l'atzar.

D'acord amb aquest procediment general, la psicometria ofereix un conjunt de teories, mètodes i tècniques vinculats al desenvolupament i l'administració de tests que donen el suport necessari als punts 2, 3 i 4 del mètode científic quan s'utilitza en el context de la mesura indirecta dels fenòmens psicològics no observables. I ho fa, com hem assenyalat anteriorment, partint d'una definició precisa dels fenòmens objecte de mesura i d'una selecció dels comportaments implicats que seran emprats com a evidències observables. Tot i això, és important fer èmfasi en algunes dificultats específiques que, com a disciplina científica, la psicometria ha d'enfrontar per tal de garantir la confiança en les inferències establertes sobre els fenòmens psicològics no observables a partir de les puntuacions obtingudes mitjançant tests. Totes elles, en íntima relació amb el repte essencial de la psicometria, són conseqüència directa de la dificultat afegida que representa la impossibilitat d'observar i manipular directament els fenòmens psicològics objecte d'interès. En aquest sentit, entre els **reptes específics** en podem assenyalat alguns dels més importants:

- **La mesura indirecta dels fenòmens psicològics no és unívoca.** Com hem anat veient, no hi ha una única manera de definir el procés de mesura així com tampoc no hi ha una única manera d'entendre els propis fenòmens objecte de mesura. Són moltes les decisions que els professionals interessats pel desenvolupament i l'administració de tests han de prendre durant aquest procés i, per tant, són moltes les possibles solucions als problemes que sorgeixen en l'aproximació indirecta a la mesura dels fenòmens psicològics. Entre aquestes decisions destaquen les relacionades amb l'elecció del que és rellevant mesurar, la selecció i el mostreig dels comportaments observables vinculats i la definició de les característiques que els tests han de tenir per a suscitar-los i mesurar-los adequadament.
- **La teoria té un paper fonamental en la mesura psicomètrica.** Totes les decisions que prenen els professionals interessats pel desenvolupament i l'administració de tests han d'estar fonamentades, com hem assenyalat anteriorment, per dos tipus de teories. D'una banda, per una teoria substantiva ben establerta entorn dels fenòmens psicològics i els comportaments de les persones que servirà de context de referència per al procés de mesura. De l'altra, una teoria dels tests, que permetrà adoptar l'enfocament més adequat per a portar a terme la mesura i establir la relació entre les puntuacions obtingudes i els fenòmens psicològics no observables objecte

d'interès. Si els tests tenen sentit i serveixen com a instruments de mesura al camp de la psicologia és, precisament, gràcies al suport que aquests dos tipus de teories proporcionen durant tot el procés.

- **La selecció i el mostreig dels comportaments imposa limitacions importants.** Un tercer repte important per a la mesura indirecta que es proposa la psicometria és la selecció i el mostreig dels comportaments emprats com a evidències observables. En consonància amb la teoria substantiva, el desenvolupament i l'administració de tests parteix d'una definició que orienta cap a uns comportaments vinculats i no cap a d'altres que, a la vegada, han de ser mostrejats adequadament atesa la impossibilitat material de suscitar i mesurar el domini sencer de comportaments al qual pertanyen. De fet, un comportament determinat pot ser emprat com a evidència de fenòmens psicològics diferents en funció de la perspectiva teòrica amb què es fonamenta el procés de construcció dels tests. Limitacions importants vinculades al temps requerit i les condicions d'administració també han de ser resoltes adequadament per a fer els tests operatius des d'un punt de vista pràctic.
- **Les puntuacions obtingudes requereixen una interpretació adequada.** Les puntuacions obtingudes mitjançant la utilització de tests no són, *per se*, informatives i han de ser sempre interpretades per a respondre al propòsit amb què els tests han estat desenvolupats. Tant des d'un punt de vista general, amb relació al contingut i el seu propòsit com amb relació al sistema de referència emprat –recordem, d'acord amb la norma, és a dir, amb relació a l'execució d'un grup de referència, o referits a un criteri, és a dir, amb relació al grau d'adequació o consecució d'aquest criteri– la utilització de tests requereix uns coneixements i unes destreses que només podem trobar a mans de professionals qualificats i sempre ha de complir amb els objectius més generals que persegueix l'avaluació psicològica a la qual serveix.
- **La mesura dels fenòmens psicològics no està lliure d'error.** L'error és un component inherent a qualsevol procés de mesura i ha de ser objecte d'un tractament adequat per a conèixer-lo i minimitzar-lo. La mesura indirecta afegeix una dificultat específica que, de manera sintètica, trobem en la impossibilitat d'observar directament tant els fenòmens psicològics objecte d'interès com l'error comès durant el procés de mesura. Partint de les puntuacions obtingudes, i gràcies al suport dels models de mesura desenvolupats en les diferents teories dels tests, la psicometria dedica una bona part dels seus esforços a aconseguir la precisió necessària que, des d'un punt de vista científic, també és exigible a la mesura indirecta dels fenòmens psicològics.
- **La mesura psicològica no es pot entendre si no és amb relació a altres mesures o esdeveniments observables.** Tot i la importància de la teoria substantiva en el desenvolupament dels tests, no podem obviar final-

ment que la mesura indirecta dels fenòmens psicològics no té sentit únicament en si mateixa sinó també amb relació a d'altres mesures derivades de la mateixa teoria o d'altres comportaments observables que corroborin el procés de mesura mitjançant tests. Com hem dit anteriorment, la mesura mitjançant tests ha de proporcionar informació rellevant amb l'objectiu de fer descripcions, classificacions, diagnòstics, explicacions o prediccions que permetin orientar una acció o la presa d'una decisió sobre el comportament de les persones, i és per això que les puntuacions obtingudes mitjançant tests requereixen un suport extern a la seva pròpia formulació per a demostrar-ne la utilitat pràctica.

Per a enfrontar-se a aquests reptes específics, la mesura indirecta dels fenòmens psicològics es proveeix d'uns **principis bàsics** que garanteixen la confiança en les inferències establertes a partir de les puntuacions obtingudes mitjançant tests. Aquests principis serveixen per a avaluar les propietats psicomètriques dels ítems i dels tests en conjunt, i entre aquests principis bàsics destaquen els següents:

- **Fiabilitat:** la precisió amb què els tests porten a terme la mesura.
- **Validesa:** la confiança en el fet que les mesures es corresponen realment amb el que es proposen mesurar.

Els diferents mètodes i procediments vinculats a aquests principis són la clau de l'èxit dels tests com a instruments d'avaluació psicològica i seran, per tant, objecte d'un tractament en profunditat en mòduls posteriors.

3. Construcció i administració de tests

Presentats els fonaments de la psicometria, en aquest darrer apartat ens farem càrrec del procés de construcció i administració de tests. Per fer-ho, començarem desenvolupant les diferents fases implicades en el disseny i construcció d'un nou test. A continuació, i partint de les claus que ens proporciona el coneixement d'aquest procés de construcció, discutirem alguns criteris importants per tal d'avaluar les característiques dels tests disponibles en la literatura i valorar-ne la conveniència amb relació als objectius de l'avaluació psicològica que han de servir. Finalment, abordarem els aspectes ètics i deontològics vinculats a la utilització de tests en el context general de la pràctica professional de la psicologia.

3.1. El procés de construcció de tests

El desenvolupament d'instruments de mesura és un procés fonamental per a qualsevol disciplina científica. I no ho és menys per a la psicometria, en què com ja hem discutit abastament la impossibilitat d'observar i manipular els fenòmens objecte d'interès afegeix una complexitat important. A continuació abordarem **deu fases fonamentals** en què, a tall de conclusió, podem organitzar el procés de construcció d'un nou test. Aquestes fases ens permeten donar una perspectiva de conjunt a aquesta introducció a la psicometria i fer una síntesi dels conceptes més importants que hem anat discutint.

Tot i que no són necessàriament seqüencials, i en tot cas no els cal seguir estrictament aquest ordre, aquestes deu fases representen algunes de les decisions més importants que els professionals interessats pel desenvolupament de nous tests han de prendre. Com veurem més endavant, conèixer aquest procés és fonamental no solament per a garantir la qualitat en el disseny i construcció de nous tests, sinó també perquè ens permet desenvolupar alguns criteris importants per a l'avaluació dels tests disponibles en la literatura. Així:

1) **Delimitació de la finalitat del test.** El desenvolupament d'un nou test comença amb una determinació clara del propòsit per al qual es vol recollir informació rellevant en el context de l'avaluació psicològica. Tal com hem assenyalat, els tests poden servir multitud de finalitats però entre elles en podem destacar algunes d'importants com són descriure, classificar, diagnosticar i explicar el comportament de les persones o fer prediccions sobre aquest. Un propòsit ben delimitat és la primera condició a complir per a l'èxit en la construcció del nou test i en permetrà l'encaix en el procés general d'avaluació psicològica.

Per saber-ne més

L'estudiant interessat pot trobar explicacions més detallades a Muñiz (1996), Muñiz i Fonseca-Pedrero (2008), Murphy i Davidshofer (2005), Downing (2006), Chadha (2009) o Rust i Gollombok (2009).

2) Definició dels fenòmens psicològics objecte de mesura. D'acord amb el procés d'inferència psicomètrica, el segon pas per a la construcció d'un test consisteix en la delimitació precisa dels fenòmens que es volen mesurar. Per a fer-ho, recordem-ho, és necessària una teoria substantiva sòlida i ben contrastada entorn dels fenòmens psicològics i els comportaments de les persones que servirà de referència en el procés de la mesura mitjançant el nou test. Aquest marc teòric és fonamental en el moment de planificar una representació adequada dels fenòmens psicològics, i permet conèixer en detall altres instruments desenvolupats prèviament en la investigació psicològica. Avaluar-ne les limitacions i els punts forts és una bona guia per a enfocar el treball necessari per a donar suport teòric al nou test.

3) Selecció i mostreig dels comportaments observables. Establert el fenomen objecte de mesura, la teoria substantiva proporciona també el context necessari per a escollir els comportaments implicats que seran emprats com a evidències observables. La seva representació adequada és fonamental per a no ometre cap comportament rellevant i també per a evitar incloure'n d'altres no directament relacionats amb l'objecte de mesura. A vegades també és recomanable portar a terme observacions, entrevistes a informadors clau o grups de discussió que proporcionin informació complementària sobre els comportaments d'interès a partir de les experiències dels participants. Com ja hem discutit, aquesta no és una fase menys importat, en tant que és la base per al desenvolupament correcte dels ítems que conformaran el nou test.

4) Especificació de les característiques del test. Delimitat l'objectiu, definit l'objecte de mesura i seleccionades les evidències observables necessàries, el pas següent consisteix a triar les característiques del nou test per a suscitar i mesurar adequadament els comportaments d'interès. En primer lloc, decidirem si es tracta d'un test d'habilitat o de potència, en què s'avaluaran l'encert i l'error en les respostes per tal de determinar la competència, l'aptitud o els coneixements dels individus, o d'una prova per a avaluar motivacions, preferències, opinions o actituds. És a dir, si es tracta d'una prova d'execució màxima o d'execució típica, com també són conegudes. D'acord amb la classificació dels tests que hem fet en aquesta introducció a la psicometria, també és el moment per a decidir el format del nou test optant per una prova de llaips i paper, de manipulació o de mesures fisiològiques. Així mateix s'han de preparar els materials necessaris i es determinarà el tipus d'administració més adequada, triant bàsicament entre una prova individual o d'administració en grup. Finalment, també es decidirà el mètode més adequat per a interpretar les puntuacions obtingudes, a partir de la norma o amb referència a un criteri.

5) Desenvolupament dels ítems que conformaran el test. Especificades les característiques generals del nou test, en aquesta fase es portarà a terme el desenvolupament dels elements que contindrà. Aquesta no és una tasca senzilla i generalment comporta la col·laboració d'un grup d'experts en el camp per a trobar la millor representació de les mostres de comportament seleccionades. De fet, no és poc habitual desenvolupar molts més ítems dels estricta-

ment necessaris per a avaluar-ne el comportament en el marc del test i seleccionar els més adequats amb relació al propòsit del nou test. L'objectiu final és disposar de tants com siguin necessaris per a representar adequadament les diferents dimensions dels fenòmens psicològics objecte d'interès. Aquest també és el moment de decidir el format que adoptaran les respostes, que serviran per a codificar els comportaments de manera estructurada i d'acord amb unes regles clares.

6) Elecció d'una teoria dels tests. Tancant el cercle establert en el procés d'inferència psicomètrica, la fase següent en el procés de construcció implica la tria d'una teoria dels tests que, mitjançant un model de mesura psicomètrica determinat, servirà per a relacionar els fenòmens psicològics objecte d'interès i les puntuacions obtingudes mitjançant els ítems que conformen el nou test. Aquesta elecció és cabdal tenint en compte les seves conseqüències en els mètodes i tècniques emprats posteriorment per a avaluar les propietats dels ítems i del test en conjunt amb l'objectiu de garantir la confiança en les inferències establertes sobre el comportament de les persones a partir de les puntuacions obtingudes.

7) Realització d'una prova pilot. Una vegada construït el test, incloent-hi els ítems potencials i el format de les respostes, és necessari redactar les instruccions que l'acompanyaran i definir les condicions en què serà administrat. Una prova pilot servirà per a avaluar el grau de comprensió d'aquestes instruccions i la viabilitat de les condicions per a administrar el test, detectar possibles dificultats en el moment de registrar les respostes i fer una primera anàlisi de les propietats, tant dels ítems com del test en el seu conjunt, d'acord amb els mètodes i tècniques indicats per la teoria dels tests utilitzada. És el moment de valorar-ne el comportament en el procés de mesura i, a partir d'aquesta informació, refinar el test modificant, descartant o afegint nous ítems. Es poden requerir proves addicionals per a avaluar correctament les modificacions introduïdes abans de tancar aquesta fase i procedir amb el desenvolupament de l'estudi de camp final.

8) Desenvolupament de l'estudi de camp. Una vegada establert el test definitiu, el pas següent consisteix a administrar-lo a la població a què es dirigeix. Per a fer-ho, se selecciona la mostra de participants necessaris que, com ja hem assenyalat, pot ser probabilística o no probabilística en funció de si volem representar la població de referència o avaluar un grup de persones que compleix unes determinades característiques. Amb aquesta informació es desenvolupen les normes o barems per a permetre la interpretació de les puntuacions amb relació a l'execució del grup de referència. En cas que es tracti d'un test referit a criteri, en lloc de normes o barems es determinen els punts de tall que permetran distingir els diferents graus d'adequació o consecució del criteri. Així mateix, s'aprofundirà en el treball sobre les propietats psicomètriques dels ítems i del test en el seu conjunt, atenent especialment a la fiabilitat i validesa de les

mesures obtingudes. Aquesta és una part fonamental del procés de construcció de tests i és la que en darrer terme garanteix que la mesura psicomètrica compleix amb totes les exigències científiques.

9) Elaboració del manual del test. Amb tota aquesta informació es desenvolupa la documentació que acompanyarà al nou test, en què s'ha d'incloure informació rellevant relativa a les diferents fases involucrades en la seva construcció: fonamentació teòrica, finalitat i població a la qual s'adreça, instruccions per a l'administració, informació per a la interpretació de les puntuacions obtingudes i anàlisi de les propietats psicomètriques. Les primeres publicacions a revistes científiques serveixen per a començar a difondre tota aquesta informació, que com ja hem comentat no sempre acaba constituint un manual pròpiament dit.

10) Revisió i millora del test. Amb la publicació del test, ja sigui difonent-lo sota llicència o lliurement, el test es posa a disposició de la comunitat científica per a obtenir noves evidències que, mitjançant el treball independent de diferents investigadors, serviran per a millorar el coneixement sobre el seu funcionament i les seves propietats psicomètriques, així com per a adaptar-lo a altres entorns socioculturals o a d'altres poblacions diferents per a les quals ha estat desenvolupat el nou test. Canvis en els fenòmens psicològics objecte de mesura, nous avenços en l'aproximació teòrica a aquests fenòmens o l'adaptació del test a noves condicions d'administració o poblacions d'interès són algunes de les raons que justifiquen un treball de revisió i actualització pràcticament indefinida per tal de refinar-ne el funcionament i valorar-ne adequadament la utilitat al servei dels objectius de l'avaluació psicològica.

3.2. Criteris per a la valoració de tests

Construir un test nou no és la pràctica més habitual en l'exercici professional de l'avaluació psicològica. De fet, com hem vist, és un treball laboriós que requereix un esforç important per part de diversos professionals durant un període de temps prolongat per a planificar, executar i analitzar les dades recollides en un o més estudis de camp. És per això que la majoria de les situacions en què s'utilitzen els tests no comencen necessàriament amb la creació d'un original. En canvi, parteixen de la cerca i valoració d'algun test existent en la literatura que encaixi amb els objectius de l'avaluació psicològica. Per tal de simplificar aquest procés, organitzacions vinculades a l'exercici professional de la psicologia com l'American Psychological Association (<http://www.apa.org>) o el Consejo General de Colegios Oficiales de Psicólogos (<http://www.cop.es>) suggereixen algunes fonts per a la cerca de tests:

- **Publicacions periòdiques.** Amb format de llibre o d'enciclopèdia, diverses publicacions es dediquen a recopilar els tests comercials disponibles amb diferents nivells de detall. Per exemple, el Buros Center for Testing de la Universitat de Nebraska edita *Tests in Print* (TIP, <http://buros.org/tests-print>) i *Mental Measurements Yearbook* (MMY,

asurements-yearbook). TIP és una de les recopilacions en llengua anglesa més exhaustives que proporciona informació rellevant sobre els objectius, les poblacions a qui s'adrecen, els autors i editorials que els publiquen i el preu de més de 3.000 tests. Per la seva banda, MMY va més enllà de l'estricta recopilació i dedica els seus esforços a proveir diferents avaluacions sobre la qualitat dels tests. Per la seva banda, una altra referència internacional important és Pro-Ed (<http://www.proedinc.com>), que edita també amb caràcter periòdic els seus *Test Publisher*, amb l'objectiu de recopilar els tests comercials disponibles, i *Test Critiques Publisher*, que funciona com un recurs complementari de l'anterior i també proporciona informació sobre la seva qualitat.

- **Bases de dades.** D'altra banda, en la Xarxa podem trobar alguns directoris oberts que proporcionen informació sobre els tests indexats, tant comercials com no comercials, i hi faciliten l'accés. Entre d'altres, podem destacar la *Test Collection* (http://www.ets.org/test_link) de l'Educational Testing Service, que inclou més de 25.000 tests. En aquest sentit, una altra font interessant és *Test Reviews Online* (<http://buros.unl.edu/buros/jsp/search.jsp>), una base de dades desenvolupada pel Buros Center for Testing que permet la cerca electrònica entre els tests incorporats a l'MMY i que ofereix accés comercial individual a les seves avaluacions.
- **Catàlegs de les editorials.** D'especial interès per a la cerca de tests en llengua no anglesa, i en especial en castellà, diferents editorials espanyoles publiquen els catàlegs de tests als quals donen llicència i proveeixen de la informació necessària per a conèixer-ne les característiques. Entre aquests, podem destacar els catàlegs de TEA Ediciones (<http://www.teaediciones.com>), del Grupo Albor-Cohs (<http://www.psicologia365.com>), de l'Instituto de Orientación Psicológica EOS (<http://www.eos.es>) i de Pearson (Pearson Clinical & Talent Assessment, <http://www.pearsonpsychcorp.es>).
- **Cerca de tests no comercials.** Finalment, hem de tenir present que no sempre els tests comercials disponibles en el mercat s'ajusten a les necessitats de l'avaluació. En aquest cas, en la cerca de tests no comercials per a utilitzar-los en l'avaluació psicològica, és convenient consultar les bases de dades que inclouen les publicacions científiques més rellevants en l'àmbit de coneixement de la psicologia. Més enllà de les recomanacions habituals, podem destacar *PsycTESTS* (<http://www.apa.org/pubs/databases/psyc-tests>), una base de dades especialitzada en tests no comercials desenvolupada per l'American Psychological Association i actualitzada amb una periodicitat mensual, en què es recopilen tests obtinguts dels mateixos autors, de publicacions en revistes científiques, llibres o tesis.

Una vegada trobat un o més tests potencialment interessants, és necessari avaluar-ne les característiques per a **valorar la conveniència del seu ús** amb relació als objectius de l'avaluació psicològica. Per a fer-ho, podem tornar al pro-

cés de construcció de tests exposat en l'apartat anterior i oferir alguns criteris a partir de les preguntes més importants que els professionals interessats en el seu ús han de formular-se amb relació als objectius del test, les seves propietats psicomètriques, el procés d'administració, el seu estatus i, finalment, l'adequació específica al procés d'avaluació:

- **Amb relació als objectius del test.** Entre d'altres, està clarament definit l'objectiu del test? Disposa del suport teòric necessari? Disposem de les evidències científiques necessàries per a utilitzar-lo com a instrument de mesura? S'actualitza regularment o, al menys, és objecte d'investigació per a valorar-ne les propietats psicomètriques i la utilitat d'acord amb els objectius que es proposa?
- **Amb relació a les seves propietats psicomètriques.** D'acord amb el procés de construcció, es proporciona la informació necessària sobre l'anàlisi dels ítems que componen el test? Quines evidències es proporcionen sobre la fiabilitat i la validesa de les mesures obtingudes mitjançant el test? Quins tipus de mostres, i amb quines grandàries, s'han utilitzat per a recollir aquestes evidències?
- **Amb relació al procés d'administració.** Disposa el test d'un manual adequat amb la informació i els materials necessaris per al seu ús? Té instruccions clares o plantilles per a la correcció? Si fos necessari, aquest manual disposa de normes o barems per a la interpretació de les puntuacions obtingudes?
- **Amb relació a l'estatus.** Es tracta d'un test comercial i, en cas de ser-ho, quin cost tenen el manual, les còpies del test, els fulls de resposta i la plantilla de correcció? És necessari algun tipus de qualificació o acreditació per a administrar-lo? Ha estat avaluat per alguna entitat o organisme independent?
- **Amb relació a l'adequació al procés d'avaluació.** Finalment, és pertinent d'acord amb els objectius d'avaluació? És el seu marc teòric congruent amb aquests objectius? Es dirigeix específicament a la població a la qual es vol administrar? Està adaptat o disposa d'una versió adaptada al context cultural en què vol ser utilitzat? S'ajusta a les condicions d'administració previstes en el procés d'avaluació, per exemple, tenint en compte el mitjà d'administració, els materials necessaris, el format de resposta als ítems o el temps requerit?

Avaluar un test abans d'utilitzar-lo és més que una pràctica convenient en l'exercici professional de la psicologia i, de fet, representa un repte important al qual les mateixes organitzacions professionals han dedicat diferents esforços per tal de sistematitzar-ne l'abordatge. Sense aturar-nos en les seves especificitats, i tenint en compte el diferent nivell de complexitat en què han estat desenvolupades, aquestes iniciatives es basen també en l'anàlisi de les de-

cisions preses en el procés de disseny i construcció dels tests, i proporcionen un marc d'avaluació de la qualitat dels tests disponibles en la literatura molt útil per a la valoració de la seva conveniència amb relació als objectius de l'avaluació psicològica.

Així, en el context internacional, hem de destacar la iniciativa de l'American Educational Research Association, l'American Psychological Association i el National Council on Measurement in Education (2014), que des de 1985 publiquen conjuntament els seus *The Standards for Educational and Psychological Testing*. D'altra banda, l'European Federation of Psychologists' Associations (2012) treballa en el desenvolupament de l'*EFPA Review model for the description and evaluation of psychological and educational tests* amb la intenció d'harmonitzar els criteris i processos d'avaluació de la qualitat dels tests a escala europea. Finalment, en el context estatal, hem de fer menció de la contribució del Consejo General de Colegios Oficiales de Psicólogos per a adaptar el model de l'EFPA a "Un modelo para evaluar la calidad de los tests utilizados en España" (Prieto i Muñiz, 2000). Aquest model ja ha estat utilitzat amb èxit per a avaluar la qualitat de deu dels tests més utilitzats pels psicòlegs espanyols (Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillo-Álvarez i Peña-Suárez, 2011).

3.3. Aspectes ètics i deontològics en l'ús de tests

Com hem discutit àmpliament al llarg d'aquesta introducció a la psicometria, l'ús de tests com a instruments d'avaluació psicològica és una de les pràctiques habituals en l'exercici de la psicologia. Tant al servei de la intervenció en els diferents camps professionals com en el cor del desenvolupament dels camps acadèmics i científics, els tests són uns instruments indispensables per a la mesura objectiva i estandarditzada dels fenòmens psicològics. I com a tals, es troben subjectes a una reflexió sobre les **conseqüències ètiques i morals que comporta el seu ús**. És a dir, en tant que instruments, poden ser utilitzats o no de manera adequada i d'acord amb unes finalitats o unes altres. No és aquest el lloc per a començar una discussió filosòfica en profunditat sobre el que significa actuar bé o fer les coses correctament, però sí per a concloure aquesta introducció reflexionant sobre algunes qüestions importants que tenen a veure amb un ús dels tests responsable, just i respectuós amb els drets i la dignitat de les persones.

Per fer-ho, repassarem alguns dels codis de conducta més importants desenvolupats en el camp de la psicologia que, en últim terme, presideixen el desenvolupament correcte de l'activitat professional, docent i investigadora. I analitzarem la manera en què aborden els aspectes ètics i morals vinculats a l'ús de tests en el context general de l'avaluació psicològica. Aquests codis són la resposta específica que la psicologia ha donat a un debat complex que, com d'altres disciplines, ha afrontat a partir de la Segona Guerra Mundial. Partint del Codi de Nuremberg (1947), la Declaració d'Hèlsinki (1964), l'Informe Belmont (1979) i les directrius proposades pel Council for International Organi-

⁽¹³⁾Per a una interessant discussió sobre els orígens podeu veure Israel i Hay (2006).

sations of Medical Sciences (1982), la consideració de les implicacions ètiques i morals del treball amb humans ha format i forma part de la pràctica quotidiana de professionals i investigadors¹³. En l'àmbit concret de la psicologia, són molts els codis que les organitzacions professionals dels diferents països han desenvolupat, però en aquest text ens referirem als treballs de l'American Psychological Association, l'European Federation of Psychologists' Associations i, en el cas espanyol, el Consejo General de Colegios Oficiales de Psicólogos. Finalment, comentarem unes directrius internacionals que, en referència al cas específic de l'ús dels tests, ha desenvolupat la International Test Commission.

Com una de les referències més importants a escala internacional, l'American Psychological Association ha contribuït al debat sobre les implicacions ètiques i morals en el camp de la psicologia declarant un dels primers codis deontològics. A partir dels anys cinquanta, i en revisions successives fins a l'actualitat, aquesta associació promou el que anomena *Ethical principles of psychologists and code of conduct* (American Psychological Association, 2010), en què estableix un preàmbul, cinc principis generals i deu estàndards que estableixen els límits i les pautes adequades per a una pràctica professional responsable. El preàmbul i els cinc principis es presenten com un posicionament ètic i moral, és a dir, un compromís amb els ideals de la psicologia i no tant com unes normes de compliment obligat. En aquest sentit, el codi reclama l'aspiració a la beneficència i no maleficència, la fidelitat i responsabilitat, la integritat, la justícia i el respecte pels drets i la dignitat de les persones.

En altres paraules, els psicòlegs han de fer tot el possible per a cercar i preservar el benestar de les persones, evitant produir cap mal i vetllant pels drets de les persones amb les quals interactuen en els diferents àmbits d'activitat professional. Així mateix, han de desenvolupar relacions de confiança, responsabilitzant-se del seu comportament i tractant de resoldre activament els conflictes d'interès que puguin presentar-se en l'exercici professional. A més a més, els psicòlegs han d'actuar honestament i amb rigor, evitant en tot moment l'engany o l'equívoc. I han de fer-ho també amb equitat i justícia, fent una atenció especial a assegurar que cap baux ni límit en la seva competència o en la seva experiència professional contribueixi o doni suport a pràctiques injustes. Finalment, els psicòlegs han de vetllar pels drets de les persones a la privacitat, confidencialitat i dignitat, respectant les diferències culturals i individuals, i evitant contribuir o donar suport a qualsevol tipus de prejudici en les seves pràctiques professionals.

Per la seva banda, els estàndards es presenten com un conjunt de normes de compliment obligat per als membres de l'associació i els sotmet a les decisions o sancions del seu comitè ètic en cas de violació del codi en l'exercici professional. Aquests estàndards estan descrits de manera general i amb diferent nivell d'articulació, amb l'objectiu que puguin ser aplicats de manera transversal als diferents contextos d'activitat professional, docent o investigadora. Els estàndards remetent a deu aspectes fonamentals com són la resolució dels conflictes ètics, la competència professional, les relacions humanes, la privacitat i la

Per saber-ne més

L'estudiant interessat pot trobar a Campbell, Vasquez, Behne i Kinscherff (2010) un comentari detallat d'aquests estàndards i també uns dilemes a tall d'exemple de la seva aplicació en la pràctica.

confidencialitat, la publicitat i les contribucions públiques, el manteniment de la documentació professional i els honoraris, la formació, la investigació i la publicació dels resultats, l'avaluació i la teràpia.

En aquest sentit, i d'acord amb els objectius d'aquesta discussió sobre els aspectes ètics i deontològics vinculats a l'ús de tests, l'estàndard relatiu a l'avaluació descriu amb detall el marc general en què aquesta ha de ser portada a terme i planteja la necessitat de basar qualsevol judici o conclusió en les evidències obtingudes mitjançant l'ús apropiat dels instruments d'avaluació. A més a més, fa referència explícita a l'ús de tests i planteja la necessitat d'utilitzar tests vàlids i fiables, l'obligació de posar a disposició de les persones les seves respostes i la seva interpretació de manera confidencial, el desenvolupament de nous tests sempre d'acord amb els estàndards de qualitat de la psicometria, la desestimació de tests obsolets o no adequats per a la població objectiu i el compliment de les condicions i ajust a les llicències d'ús dels tests desenvolupats per tercers.

Per la seva banda, l'European Federation of Psychologists' Associations (2005) ha proposat un *Meta-code of ethics* que serveix com a marc de referència per al context europeu i cerca homogeneïtzar el tractament de les qüestions ètiques i deontològiques de les diferents organitzacions professionals dels estats membres. Per a fer-ho, aquest codi conté un preàmbul general en què estableix la seva obligació de vetllar pel desenvolupament professional dels psicòlegs europeus en totes aquestes qüestions i proposa quatre principis fonamentals que, articulats en un conjunt de recomanacions, han de regir als diferents codis, procediments i comitès ètics nacionals.

Així, el codi europeu planteja la defensa i el respecte dels drets i la dignitat de les persones, el reconeixement de la importància de la competència dels professionals que exerceixen la psicologia, la responsabilitat envers les persones, les comunitats i la societat en general i la promoció de la integritat en el desenvolupament de l'activitat professional, docent i investigadora. Tant els principis com les diferents especificacions es consideren interdependents, de manera que es planteja obertament la necessitat d'establir un debat i diàleg professional al davant de la complexitat dels conflictes ètics que es presenten en la pràctica. El contingut del seu articulat reclama l'atenció dels psicòlegs per qüestions importants relatives a l'exercici professional quotidià comparables a les establertes en el codi de l'American Psychological Association però, a diferència d'aquest, no fa cap referència explícita als processos d'avaluació psicològica ni a l'ús de tests en aquest context.

D'altra banda, amb relació a l'exercici professional en el camp de la psicologia a Espanya, el Consejo General de Colegios Oficiales de Psicólogos (2010) ha desenvolupat el seu *Codi deontològic* com la plasmació d'uns drets i deures professionals que, en últim terme, serveixen de base per al judici de la conducta dels seus col·legiats (Bermejo, 2009, per a una discussió de les modificacions més recents). Disposa d'un títol preliminar, a mode de preàmbul, que en

determina l'abast i emfatitza tant el respecte fonamental pel marc normatiu i jurídic com l'obligació de rebutjar qualsevol tipus d'impediment o limitació en l'exercici professional lliure i independent. També proposa uns principis generals que, a continuació, desenvolupa en un conjunt d'àrees fonamentals, i un annex final recull el reglament de la comissió deontològica estatal encarregada de vetllar per la correcta interpretació i aplicació del codi. Així, entre els principis generals que han de regir l'activitat professional dels psicòlegs a l'Estat espanyol planteja:

- 1) L'atenció primordial al benestar, la salut, la qualitat de vida i la plenitud del desenvolupament de les persones.
- 2) La protecció dels drets humans i la responsabilitat en la fonamentació objectiva i científica de les intervencions professionals.
- 3) El no-desenvolupament de pràctiques que atemptin contra la llibertat i la integritat física i psíquica de les persones o la contribució a aquestes pràctiques.
- 4) L'obligatorietat d'informar almenys els organismes col·legials si es coneix qualsevol violació dels drets humans o tractament degradant.
- 5) El respecte pels criteris morals i religiosos de les persones.
- 6) La no-discriminació per qualsevol diferència o motiu.
- 7) La denegació de qualsevol benefici o profit que es pugui treure com a conseqüència de les relacions de poder o superioritat establertes.
- 8) L'ús d'un llenguatge prudent i crític davant d'etiquetes menyspreadores o discriminatòries en els seus informes.
- 9) El respecte per l'activitat dels altres professionals i la lliure competència.
- 10) La denúncia de pràctiques il·legítimes o intrusives.
- 11) La imparcialitat davant dels possibles conflictes d'interès entre els psicòlegs o les institucions en què desenvolupen la seva activitat i els interessos de les persones.

L'articulat desenvolupa aquests principis en set àrees fonamentals com són la competència professional i la relació amb altres professionals, la intervenció, la investigació i la docència, l'obtenció i ús de la informació, la publicitat, els honoraris i la remuneració, i les garanties processals. Articulades amb diferents nivells de detall, aquestes àrees no preveuen específicament l'avaluació psicològica ni fan cap referència explícita a l'ús de tests. Aquestes qüestions queden relegades a un tractament genèric en les àrees de la competència professional i de l'obtenció i ús de la informació, de manera que es planteja la necessitat

de disposar dels coneixements necessaris per a l'ús dels mètodes, instruments, tècniques i procediments, sempre d'acord amb les evidències científiques necessàries que en garanteixin un ús adequat. Amb relació a l'obtenció i ús de la informació, circumscriu l'ús d'aquests mètodes, instruments, tècniques i procediments d'acord amb les condicions de confidencialitat i secret professional, i reconeix el dret de les persones a conèixer els seus resultats.

Finalment, més enllà dels codis deontològics professionals, hem de fer menció del treball específic que la International Test Commission (2000) ha desenvolupat en les seves *International guidelines for test use*. Atesa la disparitat en la normativa i el nivell de desenvolupament dels codis professionals dels diferents països, aquesta organització internacional planteja unes directrius específiques sobre l'ús dels tests i el seu paper en el context de l'avaluació psicològica que, entre d'altres, subscriu el Consejo General de Colegios Oficiales de Psicólogos. Tal com assenyala en la introducció, la intenció d'aquestes directrius no és afegir un altre conjunt de recomanacions a les que ja hi ha, sinó que persegueix l'objectiu de crear una estructura coherent sota la qual es puguin entendre i aplicar els diferents codis i estàndards nacionals que desenvolupen els aspectes ètics i deontològics vinculats a l'ús de tests. Per fer-ho, estructura el seu articulat a partir d'un propòsit general que estableix un ús apropiat, professional i ètic dels tests, amb respecte pels drets de les persones, les raons per les quals s'utilitzen i el context en què s'apliquen.

Així mateix, com a gran eix conductor de les seves directrius, proposa la consecució d'aquest propòsit general mitjançant el desenvolupament i l'adquisició de les competències necessàries per a portar a terme l'administració de tests, la interpretació i comunicació adequades dels resultats, i la resolució de les dificultats, malentesos i conflictes que es puguin produir durant el procés. D'aquesta manera, aquestes directrius estableixen que els usuaris competents dels tests haurien de:

1) Responsabilitzar-se de l'ús ètic dels tests:

- Actuant de manera professional i ètica.
- Assegurant-se que tenen els coneixements i les habilitats necessàries.
- Fent-se responsables del seu ús dels tests.
- Mantenint de manera segura els materials dels tests.
- Garantint la confidencialitat dels resultats.

2) Comprometre's amb les bones pràctiques en l'ús dels tests:

- Estudiant la utilitat dels tests en els processos d'avaluació.
- Triant tests ben fonamentats i apropiats per a la situació.
- Atenent a les qüestions relacionades amb l'equitat.
- Preparant adequadament les condicions d'administració.
- Utilitzant correctament els tests.
- Obtenint les puntuacions i analitzant-les amb exactitud.

- Interpretant els resultats correctament.
- Comunicant els resultats de manera clara i exacta.
- Avaluant adequadament el funcionament i les propietats dels tests.

Totes aquestes directrius estan desenvolupades, al seu torn, en un conjunt de recomanacions concretes i representen un retrat complex de les dificultats i els reptes ètics i morals que planteja específicament l'ús dels tests com a instruments d'avaluació psicològica. Finalment, el text conclou amb uns annexos que desenvolupen algunes indicacions per al tractament específic de quatre qüestions importants com són el desenvolupament de polítiques i normatives sobre l'ús de tests en les organitzacions i empreses, la redacció de contractes entre les diferents parts involucrades en l'ús dels tests, l'administració de tests a persones amb discapacitats i la traducció de les pròpies directrius als diferents idiomes per part de les organitzacions estatals vinculades a l'exercici professional de la psicologia.

Com hem pogut fer notar, les qüestions ètiques i morals formen part intrínseca de l'exercici professional de la psicologia. Així ho reflecteixen els diferents codis deontològics que regeixen l'activitat dels psicòlegs als Estats Units, Europa i Espanya, mitjançant els quals es reconeix el repte que representa afrontar de manera activa, responsable i compromesa la complexitat de les situacions i contextos en què els professionals intervenen de manera quotidiana. Per aquesta raó és obligació dels psicòlegs exigir-se, i exigir als altres professionals amb qui col·laboren, un comportament exemplar davant dels dilemes ètics i morals que la pràctica pugui portar. Els tests, instruments imprescindibles per a l'avaluació i la intervenció psicològiques, no es troben al marge d'aquesta discussió malgrat que, com hem comentat, els codis no sempre fan un tractament específic sobre els processos d'avaluació i, concretament, sobre el paper que tenen els tests.

Com qualsevol altre tipus d'eina o instrument els tests poden ser emprats al servei d'unes finalitats o unes altres, però el compromís ha de ser la cerca constant i inequívoca del benefici i no perjudici de les persones, reaccionant davant de qualsevol tipus d'impediment que pugui limitar l'exercici professional honest, lliure i independent. En definitiva, respectant el dret de les persones a ser tractades i avaluades amb justícia, equitat i responsabilitat, i a conèixer els resultats i les conseqüències que de l'avaluació psicològica es puguin derivar. Com hem pogut veure, aquesta responsabilitat no es limita a l'administració de tests amb les garanties científiques necessàries sinó que també implica el domini de les teories, mètodes i tècniques que ha desenvolupat la psicometria per a comprendre'n el funcionament, valorar-ne la conveniència amb relació als objectius de l'avaluació i, finalment, evitar qualsevol tipus de perjudici que un ús inadequat pogués ocasionar. És per això que els psicòlegs han de ser curiosos amb la seva pròpia competència, la clau de volta per a un ús adequat dels

tests, amb l'objectiu de desenvolupar els coneixements i les habilitats necessàries que garanteixin un respecte profund pels drets i la dignitat de les persones a qui administren els seus tests.

Bibliografia

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, D. C.: American Psychological Association.

American Psychological Association (2010). Ethical principles of psychologists and code of conduct. Washington, D. C.: American Psychological Association. <<http://www.apa.org/ethics/code/principles.pdf>>

Anastasi, A. (1988). *Psychological testing* (6a. ed.). Nova York: Macmillan.

Bermejo, V. (2009). Motivaciones para la revisión y cambios en el código deontológico de la profesión de la psicología. *Papeles del Psicólogo*, 30 (3), 195-206. <<http://www.papelesdelpsicologo.es/pdf/1748.pdf>>

Bock, R. D. (1997). A brief history of Item response theory. *Educational Measurement: Issues and Practice*, 16 (4), 21-33.

Bock, R. D. i Jones, L. V. (1968). *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day.

Bondy, M. (1974). Psychiatric antecedents of psychological testing (before Binet). *Journal of the History of the Behavioral Sciences*, 10 (2), 180-194.

Boring, E. G. (1978). *Historia de la psicología experimental*. Mèxic: Trillas.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in modern psychometrics*. Cambridge: Cambridge University Press.

Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist*, 44(3), 576-578.

Brennan, R. L. (1997). A perspective on the history of Generalizability Theory. *Educational Measurement: Issues and Practice*, 16 (4), 14-20.

Bridgman, P. W. (1927). *The logic of modern physics*. Nova York: Macmillan.

Buchanan, R. D. i Finch, S. J. (2005). History of psychometrics. A B. S. Everitt i D. Howell (Eds.). *Encyclopedia of statistics in behavioral science* (pp. 875-878). Chichester: Wiley.

Campbell, L., Vasquez, M., Behne, S., i Kinscherff, R. (2010). *APA Ethics code commentary and case illustrations*. Washington, DC: American Psychological Association.

Campbell, N. R. (1920). *Physics: The elements*. Cambridge: Cambridge University Press.

Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. Londres: Longmans, Green, and co.

Carson, J. (1993). Army alpha, army brass, and the search for army intelligence. *Isis*, 84 (2), 278-309.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40 (3), 153-193.

Chadha, N. K. (2009). *Applied psychometry*. Nova Delhi: Sage.

Consejo General de Colegios Oficiales de Psicólogos (2010). *Código deontológico*. Madrid: Consejo General de Colegios Oficiales de Psicólogos. <<http://www.cop.es/pdf/Codigo-Deontologico-Consejo-Adaptacion-Ley-Omnibus.pdf>>

Crocker, L. i Algina, J. (2006). *Introduction to classical and modern test theory*. Mason: CENGAGE Learning.

Cronbach, L. J., Rajaratnam, N., i Gleser, G. C. (1963). Theory of Generalizability: A liberation of reliability theory. *British Journal of Mathematical and Statistical Psychology*, 16 (2), 137-163.

Dingle, H. (1950). A theory of measurement. *The British Journal for the Philosophy of Science*, 1(1), 5-26.

Downing, S. M. (2006). Twelve steps for effective test development. A S. M. Downing i T. M. Haladyna (Eds.). *Handbook of test development* (pp. 3-25). Mahwah: Lawrence Erlbaum.

DuBois, P. H. (1970). *The history of psychological testing*. Boston: Allyn & Bacon.

European Federation of Psychologists' Associations (2005). *Meta-code of ethics*. Brussel·les: European Federation of Psychologists' Associations. <<http://www.efpa.eu/ethics/ethical-codes>>

European Federation of Psychologists' Associations (2012). *EFPA Review model for the description and evaluation of psychological and educational tests. Version 4.2.4*. Brussel·les: European Federation of Psychologists' Associations. <<http://www.efpa.eu/professional-development>>

Ferguson, A., Myers, C. S., Bartlett, R. J., Banister, H., Bartlett, F. C., Brown, W., Campbell, N. R., Craik, K. J. W., Drever, J., Guild, J., Houstoun, R. A., Irwin, J. O., Kaye, G. W. C., Philpott, S. J. F., Richardson, L. F., Shaxby, J. H., Smith, T., Thouless, R. H., i Tucker, W. S. (1940). Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *Advancement of Science*, 1, 331-349.

Fernández-Ballesteros, R. (1997). Evaluación psicológica y tests. A A. Cordero (Ed.). *La evaluación psicológica en el año 2000* (pp. 11-26). Madrid: TEA Ediciones.

Fraser, C. O. (1980). Measurement in psychology. *British Journal of Psychology*, 71(1) 23-34.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *History of Psychology*, 87 (3), 564-567.

García Cueto, E. (1993). *Introducción a la psicometría*. Madrid: Siglo Veintiuno de España Editores.

Gibby, R. E., i Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11(3), 164-184.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 18 (8), 519-521.

Gleser, G. C., Cronbach, L. J., i Rajaratnam, N. (1965). Generalizability of scores influenced by multiple sources of variance. *Psychometrika*, 30 (4), 395-418.

Goodenough, F. L. (1949). *Mental testing: Its history, principles, and applications*. Nova York: Rinehart.

Goslin, D. A. (1963). *The search for ability. Standardized testing in social perspective*. Nova York: Russell Sage Foundation.

Gruijter, D. N. M. de i Kamp, L. J. T. van der (2008). *Statistical test theory for the Behavioral Sciences*. Boca Raton: Chapman & Hall.

Guilford, J. P. (1936). *Psychometric methods*. Nova York: McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests*. Nova York: Willey.

Gulliksen, H. (1968). Louis Leon Thurstone, experimental and mathematical psychologist. *History of Psychology*, 23(11), 786-802.

Hambleton, R. K. (1994). The rise and fall of criterion-referenced measurement? *Educational Measurement: Issues and Practice*, 13 (4), 21-26.

Hambleton, R. K., Swaminathan, H., i Rogers, H. J. (1991). *Fundamentals of Item response theory*. Newbury Park: Sage Publications.

Hand, D. J. (1996). Statistics and the theory of measurement. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159 (3), 445-492.

Hilgard, E. R. (1987). *Psychology in America: A historical survey*. Orlando: Harcourt.

Holden, R. R. (2000). Psychometrics. A A. E. Kazdin (Ed.). *Encyclopedia of psychology* (vol. VI, pp. 417-419). Nova York: Oxford University Press.

International Test Commission (2000). *International guidelines for test use*. <<http://www.intestcom.org/upload/sitefiles/41.pdf>>

- Kline, P. (1998). *The new psychometrics. Science, psychology, and measurement*. Londres: Routledge.
- Jáñez, L. (1989). *Fundamentos de psicología matemática*. Madrid: Pirámide.
- Jensen, A. R. (1980). *Bias in mental testing*. Nova York: Free Press.
- Jones, L. V. (1971). The nature of measurement. A R. L. Thorndike (Ed.). *Educational measurement* (2a. ed.) (pp. 335-355). Washington, DC: American Council on Education.
- Jones, L. V. i Thissen, D. (2007). A history and overview of Psychometrics. A C. R. Rao i S. Sinharay (Eds.). *Handbook of statistics 26. Psychometrics* (pp. 1-27). Amsterdam: Elsevier.
- Kaplan, R. M. i Saccuzzo, D. P. (2001). *Psychological testing* (6a. ed.). Belmont: Wadsworth Publishing Company.
- Lord, F. M. (1953). On the statistical treatment of football numbers. *The American Psychologist*, 8(12), 750-751.
- Lord, F. M. i Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison Wesley.
- Martin, O. (1997). La mesure en psychologie de Binet à Thurstone, 1900-1930. *Revue de synthèse*, 118(4), 457-493.
- Matínez Arias, M. R., Hernández Lloreda, M. J., i Hernández Lloreda, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Maydeu-Olivares, A. i McArdle, J. J. (2005). *Contemporary psychometrics*. Mahwah: Lawrence Erlbaum.
- Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *History of Psychology*, 100(3), 398-407.
- Michell, J. (1999). *Measurement in psychology*. Cambridge: Cambridge University Press.
- Muñiz, J. (1996). *Psicometría*. Madrid: Editorial Universitat.
- Muñiz, J. (2003). *Teoría clásica de los tests*. Madrid: Pirámide.
- Muñiz, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31 (1), 57-66. <<http://www.papelesdelpsicologo.es/pdf/1796.pdf>>
- Muñiz, J., Fernández-Hermida, J. R., Fonseca-Pedrero, E., Campillo-Álvarez, A., i Peña-Suárez, E. (2011). Evaluación de tests editados en España. *Papeles del Psicólogo*, 32 (2), 113-128. <<http://www.papelesdelpsicologo.es/pdf/1947.pdf>>
- Muñiz, J. i Fonseca-Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *Revista de Investigación en Educación*, 5, 13-25.
- Muñiz, J. i Hambleton, R. K. (1992). Medio siglo de teoría de respuesta a los ítems. *Anuario de Psicología*, (52), 41-66.
- Murphy, K. R. i Davidshofer, C. O. (2005). *Psychological testing* (6a. ed.). Upper Saddle River: Pearson education.
- Nicolas, S. i Ferrand, L. (2002). Alfred Binet and higher education. *History of Psychology*, 5 (3), 264-283.
- Padilla, M., Merino, J. M., Rodríguez-Miñón, P., i Moreno, E. (1996). *Psicología matemática I*. Madrid: UNED.
- Popham, W. J. i Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6 (1), 1-9.
- Prieto, G. i Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75. <<http://www.papelesdelpsicologo.es/vernumero.asp?id=1102>>
- Rogers, T. B. (1995). *The psychological testing enterprise*. Pacific Grove, CA: Brooks / Cole Publishing Company.

- Rozeboom, W. W. (1966). Scaling theory and the nature of measurement. *Psychometrika*, 16(2), 170-233.
- Rust, J. i Golombok, S. (2009). *Modern psychometrics. The science of psychological assessment* (3a. ed.). Londres: Routledge.
- Samejima, F. (2000). Psychometric society. A. A. E. Kazdin (Ed.). *Encyclopedia of psychology* (vol. VI, pp. 419-420). Nova York: Oxford University Press.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 667-680.
- Sokal, M. M. (1982). James McKeen Cattell and the failure of anthropometric mental testing, 1890-1901. A. W. R. Woodward i M. G. Ash (Eds.), *The problematic science. Psychology in nineteenth-century thought* (pp. 322-345). Nova York: Praeger Publishers.
- Suppes, P. (1951). A set of independent axioms for extensive quantities. *Portugaliae Mathematica*, 10, 163-172.
- Swistak, P. (1990). Paradigms of measurement. *Theory and Decision*, 29 (1), 1-17.
- Thurstone, L. L. (1931). *The reliability and validity of tests: Derivation and interpretation of fundamental formulae concerned with reliability and validity of tests and illustrative problems*. Ann Arbor: Edwards Brothers.
- Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vectors of the mind*. Chicago: Chicago University Press.
- Traub, R. E. (1997). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice*, 16 (4), 8-14.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Nova York: Wiley.
- Urbina, S. (2004). *Essentials of psychological testing*. Hoboken: Wiley.
- Valentine, E. R. (1999). The founding of the psychological laboratory, University College London: "Dear Galton...Yours truly, J. Sully." *History of Psychology*, 2 (3), 204-218.
- Velleman, P. F. i Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician*, 47 (1), 65-72.
- Wolf, T. H. (1969). The emergence of Binet's conception and measurement of intelligence: A case history of the creative process. *Journal of the History of the Behavioral Sciences*, 5 (2), 113-134.
- Yela, M. (1984). *Introducción a la teoría de los tests*. Madrid: Universidad Complutense.
- Zeidner, J. i Drucker, A. J. (1988). *Behavioral science in the Army: A corporate history of the Army Research Institute*. Alexandria: United States Army Research Institute for the Behavioral and Social Sciences.