

PRÀCTIQUES DE PSICOMETRIA: ANÀLISI DE TESTS AMB EL PROGRAMA METRIX

Jordi Renom Pinsach
María Isabel Núñez Peña

CONTÉ CD

Departament de Metodologia de les Ciències del Comportament
Facultat de Psicologia



UNIVERSITAT DE BARCELONA



TEXTOS DOCENTS

330

PRÀCTIQUES DE PSICOMETRIA: ANÀLISI DE TESTS AMB EL PROGRAMA METRIX

Jordi Renom Pinsach
María Isabel Núñez Peña

Departament de Metodologia de les Ciències del Comportament
Facultat de Psicologia

Publicacions i Edicions



UNIVERSITAT DE BARCELONA



INTRODUCCIÓ

Què és aquest manual?

Aquesta obra correspon a una tercera versió dels anteriors textos-guia *Tècniques psicomètriques: anàlisi de casos* i *Anàlisi de Tests i Qüestionaris amb el programa Metrix*. El nou text no pretén ser cap manual de tècniques psicomètriques ni cap guia d'usuari d'un programa informàtic. De fet, com a text-guia es tracta bàsicament d'un document de pràctiques que combina ambdós vessants i que vol apropar-vos al procés tècnic de tractament de dades recollides amb instruments psicomètrics a nivell professional.

El manual descriu, principalment, i de manera pràctica, deu rutines d'anàlisi psicomètrica basades en l'aplicació del programa METRIX Engine en diversos fitxers de dades reals que trobareu en el CD de treball. Totes els casos o anàlisis s'apropen a situacions habituals o escenaris de treball habituals i han estat triats pensant en el tractament de les respostes a tests i qüestionaris psicomètrics representatius. Cada cas proporciona determinats resultats i demana una seqüència de decisions prèvia. Tot això permet reflexionar sobre les repercussions que té l'anàlisi sobre els examinands o individus dels quals s'han de prendre decisions.

Queda traçat, doncs, el lligam entre el contingut del CD, el programa d'anàlisi i la finalitat didàctica de les sessions de pràctiques. Globalment, l'objectiu d'aquesta combinació és facilitar la comprensió d'un procés que sovint queda deslligat de la realitat si només es tracta des del vessant teòric.

Per què s'ha creat?

El desgast i divulgació progressius dels tests, la necessitat de nous tests i el reajustament dels que ja funcionen requereixen una atenció i una formació constant per part dels seus usuaris. Mantenir la qualitat d'aquests instruments és l'única via que permet confiar en la seva utilitat i objectivitat. Les garanties dels resultats d'un examen, una exploració psicològica, una enquesta o proves de selecció depenen en bona part d'aquest manteniment. Tradicionalment, s'ha atribuït als tests una sèrie de propietats, relacionades amb la fiabilitat i la validesa, que els usuaris han acceptat. Actualment, però, cal verificar cada vegada més aquestes propietats i actualitzar-les en el propi entorn de treball per a la mostra d'individus i els objectius particulars de l'avaluació. En aquest sentit, una iniciativa per cobrir aquesta necessitat ha estat la creació d'una "Comisión de Tests" des del Colegio Oficial de Psicólogos que vetlla pel manteniment i la millora de les qualitats dels tests editats a l'Estat espanyol.

Per tant, és necessària una formació teòrica i pràctica basada en les tècniques que permeten assolir, amb autonomia, tots aquests objectius. Aquesta preparació, ha de proporcionar recursos suficients al futur professional per quan sospiti de l'eficiència de les proves en l'avaluació psicològica i educativa. Dubtes com els següents són habituals quan es treballa amb proves:

És fiable el test que apliquem? Sovint un test no proporciona prou garanties de fiabilitat i les seves puntuacions arrossegueu un error de mesura considerable que repercutirà en les nostres decisions.

Podriem millorar l'eficiència del test? Un estudi adequat pot modificar el contingut de la prova, o retallar-ne la llargada, i mantenir, i fins i tot, millorar, les seves qualitats. El temps i l'esforç dedicats a la correcció i la interpretació en sortiran beneficiats com també la reducció de la fatiga dels examinands.

És possible elaborar barems actuals i ajustats al nostre personal? Amb massa freqüència els barems i les taules de puntuacions normatives d'un test no inclouen el perfil personal amb què treballem o bé estan desfasats. La dependència d'uns barems poc representatius redueix la utilitat de les mesures obtingudes.

El treball pràctic és la millor via per respondre preguntes com aquestes i les sessions proposades en el manual, l'eina per fer-ho.

A qui va dirigit?

Principalment, als alumnes de l'assignatura *Psicometria* i d'altres optatives i de lliure elecció relacionades amb l'elaboració de tests i qüestionaris amb eines informàtiques, de la llicenciatura de psicologia de la Universitat de Barcelona. D'altra banda, hi ha una sèrie de capítols que connecten amb assignatures, com ara *Psicologia de l'activitat física i de l'esport*, *Tècniques d'investigació psicosocial*, *Tècniques d'ergonomia*, *Avaluació, selecció i formació*. També serà útil als estudiants d'altres disciplines relacionades amb la creació i l'anàlisi d'instruments de mesura i avaluació en ciències de la salut, socials i de l'educació. Igualment, els professionals i docents que disposin de dades i desitgin actualitzar les seves proves, trobaran aquí unes quantes seqüències metodològiques i tècniques de treball per aconseguir-ho.

Per a aquells que vulguin aprofundir més i treballin especialment en la selecció de personal i avaluació del rendiment, des de l'àmbit més local fins a projectes a gran escala, existeix un altre tutorial multimedia desenvolupat per la Universitat de Barcelona que amplia aquest manual.

PIERT: tutorial multimedia para el diseño de pruebas de rendimiento.

ISBN 84-8338-321-7 versió professional

ISBN 84-8338-283-0 versió aprenentatge

J. Renom Pinsach (coord.) A. Solanas Pérez, E. Doval Diéguez y M. I. Núñez Peña.

(existeix una llicència de Campus a la Universitat de Barcelona)

En general, les tasques proposades al llarg del manual només requereixen uns coneixements informàtics mínims a nivell d'usuari. Precisament, els primers capítols permeten, a qui no hagi treballat abans amb l'ordinador, assolir aquests mínims i després, a poc a poc, es van afegint més elements. Quant a la fonamentació teòrica, cal tenir les idees clares sobre els conceptes psicomètrics bàsics i anar preparant la matèria teòrica paral·lelament a les tasques d'anàlisi. Qualsevol desequilibri entre ambdues facetes comportarà dificultats de seguiment de les sessions pràctiques.

Com s'ha d'emprar?

El manual s'ha creat considerant les tasques d'anàlisi de tests com un procés integrat on es combinen els fonaments psicomètrics i l'interès per millorar els mètodes actuals d'avaluació. Per comprendre aquest procés haureu de fer un treball realment pràctic. Per tant, si us plau, no intenteu llegir sencer el manual ni entendre abans d'hora el que hi ha descrit en les anàlisis, pot ser contraproduent a l'objectiu de les pràctiques. Abans us demanem que llegiu atentament la resta de la introducció i el capítol primer.

Les pràctiques informatitzades que us presentem són de complexitat creixent i remarquen les relacions causa-efecte entre diverses decisions i els resultats finals de cada procés. A poc a poc, des de la primera sessió, us adonareu dels elements i els indicadors que informen millor de les qualitats d'una prova. Gradualment hi haurà també més autonomia de treball, fins a arribar a una plena capacitat de decisió. Emprar l'ordinador agilita situacions i decisions, ja que en poc temps es veuen les diverses solucions

possibles per a unes dades.

Quant a METRIX, les possibilitats gràfiques i de gestió (entorn WINDOWS) del programa faciliten ràpidament a l'usuari la detecció de les causes i efectes abans esmentats. Així doncs, sense una pantalla davant i la possibilitat de teclejar instruccions i comprovar resultats, poca utilitat tindrà la lectura dels capítols següents i difícilment en serà intel·ligible el contingut.

L'experiència dels darrers anys ha mostrat les condicions òptimes de treball amb el manual:

- Les sessions de pràctiques s'han de fer en grup a l'aula d'ordinadors, treballant en petits equips de dues o tres persones i amb el suport dels monitors de pràctiques. D'aquesta manera les decisions són compartides i podreu combinar millor la revisió dels conceptes teòrics amb l'enregistrament de resultats i el treball davant de la pantalla.

La consulta de la documentació de teoria durant les sessions de pràctiques s'ha de convertir en un hàbit per tal de comprendre el que us ofereixen les combinacions de resultats. Cada sessió afegeix nous elements a les anteriors. Si no enteneu la funció, els avantatges i els inconvenients de cada coeficient, gràfic o taula de resultats difícilment podreu valorar les novetats i aportacions de la sessió següent.

- És fonamental resoldre els exercicis proposats en cada anàlisi i anotar totes les observacions i dubtes que apareixin durant les sessions. Aquest és un manual de treball i podeu escriure en qualsevol lloc de les pàgines.

Què sabreu fer en acabar les pràctiques?

En acabar les pràctiques sereu capaços de planejar i executar una seqüència d'anàlisi convencional de les dades d'un test psicomètric. Sabreu valorar el potencial d'un instrument i decidir les condicions en les quals assoleix la màxima eficàcia. Si, en el futur o des d'altres assignatures, elaboreu un nou qüestionari o prova, podreu seguir el procés d'anàlisi des de zero i modelar-ne el funcionament. En qualsevol cas, dominareu les eines per actualitzar barems, verificar indicadors de qualitat i entendreu les relacions internes que regulen el funcionament d'una prova.

Tradicionalment totes aquestes tasques han estat les més àrides en el treball psicomètric, ja que no hi ha gaires eines que permetin executar les anàlisis de manera àgil. Quan arribeu a la darrera sessió d'anàlisi veureu que és el mateix haver treballat amb mostres petites d'assaig de cent examinands que amb grups grans i llargues proves o bateries. El missatge d'aquestes pràctiques és el mètode que cal seguir. Si feu els exercicis i seguïu l'ordre natural dels capítols assolireu aquesta metodologia i us serà útil en el futur.

El manual s'estructura en catorze capítols. Els cinc primers formen un bloc introductori i són molt importants per a qui no tingui cap experiència amb l'ordinador i els formats de tabulació de les dades. El capítols restants formen el nucli de treball del manual i són pròpiament les sessions d'anàlisi de casos que hem anat presentant en aquesta introducció.

ÍNDIX

CAPÍTOL I: RUTINES DE TREBALL

1. Introducció i objectius	5
2. Descripció dels capítols	5
3. Rutines d'anàlisi	7
4. Conceptes	9
4.1. Anàlisi dels ítems	9
4.2. Indicadors i qualitats del test	9
5. Resum del capítol	13

CAPÍTOL II: FORMATS DELS FITXERS DE DADES

1. Introducció i objectius	15
2. Possibilitats	15
3. El format Microcat	16
3.1. Primera línia	17
3.2. Segona línia	20
3.3. Tercera línia	22
3.4. Quarta línia	22
3.5. Cinquena línia en endavant	25
3.6. Anotacions i comentaris	26
3.7. Variables personals	26
4. El format ASCII	27
5. Resum del capítol	30

CAPÍTOL III: EXERCICIS AMB EL FORMAT MICROCAT

1. Introducció i objectius	31
2. Exercicis amb el format Microcat	31
2.1. Primer bloc: detecció d'anomalies	32
2.2. Segon bloc: elaboració de codis	34
3. Resum del capítol	37

CAPÍTOL IV: EL CD DE PRACTIQUES

1. Introducció i objectius	39
2. Contingut del CD	39
3. Els arxius amb extensió <i>.dat</i>	40
4. Els arxius amb extensió <i>.mct</i>	41
5. Els arxius amb extensió <i>.txt</i>	42
6. Els arxius amb extensió <i>.doc</i>	42
7. Resum del capítol	42

CAPÍTOL V: ANÀLISI DE *CALCUL.MCT* I ÚS DE L'AJUDA DE METRIX

1. Introducció i objectius	43
2. Procés d'importació	44
3. La fitxa del test	46
4. Procés d'anàlisi i cerca de resultats	48
5. L'ajuda de Metrix	50
6. Exercicis	51
7. Resum del capítol	52

CAPÍTOL VI: ANÀLISI DE *JEFES.MCT* I EDICIÓ DIRECTA DE DADES

1. Introducció i objectius	53
2. Procés d'importació	53
3. La fitxa del test	55
4. Procés d'anàlisi i cerca de resultats	56
5. Edició de dades des de Metrix	58
6. Exercicis	59
7. Resum del capítol	60

CAPÍTOL VII: ANÀLISI GENERAL D'*ENGLISH.MCT*

1. Introducció i objectius	61
2. Primera anàlisi	65
3. Desactivació d'ítems i segona anàlisi	68
4. Tercera anàlisi	69
5. Resum del capítol	72

CAPÍTOL VIII: ANÀLISI COMPLEMENTÀRIA D'*ENGLISH.MCT*

1. Introducció i objectius	73
2. Procediment automàtic per desactivar ítems	73
3. Anàlisi com a escalograma	75
4. Resum del capítol	80

CAPÍTOL IX: ANÀLISI GENERAL I COMPARATIVA DE *BIOTEC.MCT*

1. Introducció i objectius	81
2. Especificació de la fitxa del test	84
3. Anàlisi comparativa	85
4. Anàlisi de les alternatives	87
5. Estudis comuns i puntuacions d'examinand	98
6. Resum del capítol	90

CAPÍTOL X: ANÀLISI GENERAL D' *HALECI.MCT*

1. Introducció i objectius	91
2. Característiques d'Halec	91
3. Primera anàlisi	94
4. Segona anàlisi	95
5. El fitxer <i>halec.txt</i>	96
6. Resum del capítol	100

CAPÍTOL XI: ANÀLISI D'*HALEC2.MCT*

1. Introducció i objectius	101
2. Anàlisi global de les proves i de les dades personals	101
3. Anàlisi per torns	103
4. Anàlisi per sexes	104
5. Resum del capítol	105

CAPÍTOL XII: ANÀLISI DE *POMS.MCT*

1. Introducció i objectius	107
2. Descripció del POMS	107
3. Primera anàlisi per escales i depuració	110
4. Segona anàlisi	112
5. Comparació gràfica d'escales	113
6. Detecció de perfils	114
7. Anàlisi global	116
8. Resum del capítol	118

CAPÍTOL XIII: ANÀLISI DE *BANC.MCT*

1. Introducció i objectius	119
2. La corba d'omissió	123
3. Criteris per a la revisió dels ítems	125
4. Tipus d'ítems anòmals	127
5. Primera desactivació i segona anàlisi	128
6. Segona desactivació i tercera anàlisi	130
7. Resum del capítol	130

CAPÍTOL XIV: ANÀLISI DE *TRI.MCT*

1. Introducció i objectius	131
2. El model de Rasch	131
2.1. El perfil d'encert	132
2.2. Els models de la TRI	132
2.3. El calibratge d'ítems	135
3. Procés d'estimació de paràmetres	135
3.1. Consideracions prèvies	136
3.2. Inici del procés	136
4. Resultats	138
5. Resum del capítol	144

ANNEX: SOLUCIONS

1. Solucions capítol III	145
2. Solucions capítol V	149
3. Solucions capítol VI	150
4. Solucions capítol VII	151
5. Solucions capítol VIII	154
6. Solucions capítol IX	156
7. Solucions capítol X	157
8. Solucions capítol XI	159
9. Solucions capítol XII	160
10. Solucions capítol XIII	163
11. Solucions capítol XIV	164

CAPÍTOL I: RUTINES DE TREBALL

1. INTRODUCCIÓ I OBJECTIUS

L'objectiu d'aquest capítol és situar-vos en el procés de treball de la resta del manual. Trobareu una petita descripció de cada capítol i a continuació una sèrie de comentaris sobre el que anomenem *rutines de treball* o *d'anàlisi*. Recordeu que aquest text s'ha creat per facilitar el seguiment de deu rutines de treball habituals emprant les dades dels arxius enregistrats en el disquet de pràctiques. Heu de tenir sempre en compte la finalitat de cadascuna de les anàlisis i veure el procés de treball com un cicle que pot repetir-se fins a arribar a una situació idònia (prova, test, qüestionari) en termes de qualitat d'avaluació.

Encara que més endavant hi ha força referències de lectures de suport, en l'apartat quatre d'aquest capítol trobareu la descripció complementària dels conceptes i resultats principals tractats en les diverses anàlisis del manual. No cal que els llegiu ara, la seva finalitat és servir-vos d'ajuda durant les anàlisis quan hi hagi dubtes abans de consultar amb més profunditat la documentació de teoria.

2. DESCRIPCIÓ DELS CAPÍTOLS

Tot seguit trobareu una breu descripció dels capítols i les convencions emprades en el manual. Recordeu que els cinc primers capítols són una preparació per a les tasques d'anàlisi dels deu següents. Les sessions o casos tenen una estructura interna constant en quatre apartats:

- 1.Introducció. Cada sessió comença amb la informació següent:
 - presentació del test amb el qual es treballarà,
 - un exemplar dels ítems de la prova analitzada.
- 2.Nucli de la sessió. Seguiment d'una sèrie de punts amb diverses instruccions i decisions.
- 3.Exercicis. Els trobareu combinats amb el seguiment dels punts anteriors.
- 4.Resum del capítol. Serveix per tancar el capítol i assenyalar els elements més importants.

Al final del document hi ha l'annex amb les solucions dels exercicis de cada capítol. Amb la mateixa numeració dels exercicis trobareu la resposta correcta i el procediment per arribar-hi.

En totes les sessions en què trebal·leu amb l'ordinador veureu que les instruccions i ordres per clicar o marcar s'han escrit en **negreta**. També trobareu força vegades, quan hi hagi una decisió o punt crític, l'avís **Compte!**. Quant als noms dels diversos fitxers de dades del disquet de pràctiques, sempre s'escriuran en lletra *itàlica*.

Les sessions s'han desenvolupat sobre l'entorn de les diverses versions de WINDOWS i METRIX, i s'han respectat els termes originals en castellà dels menús i els diàlegs. En termes generals no us heu de

preocupar, ja que les instruccions són gairebé idèntiques. Quant a METRIX Engine, és un programa d'anàlisi psicomètrica preparat per acceptar una gran diversitat de situacions i proves. Hi ha dues versions de METRIX: la professional i la d'aprenentatge. Ambdues s'han escrit en castellà.

La versió professional està instal·lada a les aules d'informàtica de la facultat de Psicologia de la Universitat de Barcelona. Permet analitzar simultàniament fins a 10 proves, 250 ítems, 60.000 examinands i 26 variables personals externes.

Metrix Engine UB: Analizador de tests y cuestionarios (versión Prof.)
ISBN: 84-8338-428 Autor: Jordi Renom Pinsach (coord.)

La versió d'aprenentatge es distribueix dins del Tutorial PIERT, admet fins a 5 proves, 100 ítems, 100 examinands i 5 variables personals. Al marge d'aquestes diferències, ambdues versions ofereixen les mateixes possibilitats d'anàlisi i consulta de resultats. A ple rendiment, el límit per a la versió professional se situa al voltant de les 900 finestres de resultats. Com que és una aplicació desenvolupada sobre WINDOWS, la gestió de les finestres i el ratolí tindran un paper important en totes les sessions.

És important que us deixeu emportar per la seqüència natural de cada capítol. No llegiu abans d'hora les solucions dels exercicis ni us avanceu en la lectura del apartat d'una sessió. Només quan no trobeu la solució, i com a darrer recurs, serà aconsellable consultar els resultats i sobretot la manera com s'hi han arribat. El seguiment del manual no us ha de portar a considerar les anàlisis com una sèrie de tasques automatitzades. Cada sessió té uns objectius que s'han d'integrar amb les lectures recomanades i amb la finalitat general de millorar els instruments de mesura.

Objectius dels capítols

Capítol II. Conèixer els diferents formats o preparació dels fitxers de dades que després s'analitzaran. En aquest capítol no hi ha cap exercici ni cal ordinador.

Capítol III. Es dedica exclusivament a practicar sobre el paper les diverses possibilitats del format Microcat. Hi ha exercicis de reconeixement d'anomalies i de preparació de codificacions. L'ordinador tampoc és necessari.

Capítol IV. Aquest capítol descriu els diferents tipus d'arxius enregistrats en el disquet de pràctiques.

Capítol V. Anàlisi de les dades d'una prova de càlcul de resposta oberta (l'arxiu *calcul.mct*). És el primer procés d'anàlisi basat en un model clàssic: l'escalograma de Guttman. L'objectiu és organitzar la seqüència d'importació de dades, la configuració de l'anàlisi, la cerca de resultats i l'ús de l'ajuda de METRIX. La interpretació dels indicadors i els resultats encara no és el més important.

Capítol VI. Repeteix bona part del contingut del capítol anterior però amb les dades de l'arxiu *jefes.mct*, que correspon a un altre model clàssic: una escala tipus Likert. El qüestionari que es tracta en aquest capítol afecta les valoracions fetes per uns treballadors envers els seus caps. També es comença a treballar la possibilitat de tabular o afegir noves dades a l'arxiu directament des de METRIX.

Capítol VII. Anàlisi general d'una prova d'anglès en format d'alternativa múltiple. Les dades de l'arxiu *english.mct* portaran a valorar ja els resultats i els indicadors principals per prendre decisions sobre l'adequació de l'instrument.

Capítol XIII. Anàlisi complementària de les dades de l'arxiu *english.mct*. S'afegeix una nova modalitat de desactivació d'ítems i una anàlisi alternativa basada en el model de l'escalograma.

Capítol IX. Anàlisi de les respostes a un examen sobre coneixements bàsics de biotecnologia. Es tracta d'emprar un recurs que permet analitzar una mateixa prova de maneres diferents. Per aconseguir-ho, cal primer duplicar diverses vegades les dades dins del mateix arxiu, en aquest cas *biotec.mct*, simulant diverses proves que realment són repetició de la mateixa.

Capítol X. Planteja l'anàlisi completa de dues proves que tenen format diferent. Una mesura els hàbits de lectura i l'altra l'actitud envers la lectura com a activitat de lleure. Les dades són en un mateix arxiu anomenat *halec1.mct*.

Capítol XI. És una sessió complementària de l'anterior en la qual es treballa amb l'arxiu *halec2.mct*. Repeteix els recurs de còpia de les dades i diverses anàlisis de la mateixa prova. A més, per primer cop, s'incorporen dades personals dels examinands de l'arxiu *halec.txt*.

Capítol XII. És l'anàlisi d'una prova clàssica multidimensional de mesura de l'estat d'ànim: el POMS (Profile of Mood States). Les dades de l'arxiu *poms.mct* permetran valorar el funcionament de l'instrument amb les respostes d'un grup d'esportistes abans d'una competició. Com a novetat hi ha també un arxiu, el *pomsg.mct*, que ens permetrà analitzar la prova com si fos unidimensional.

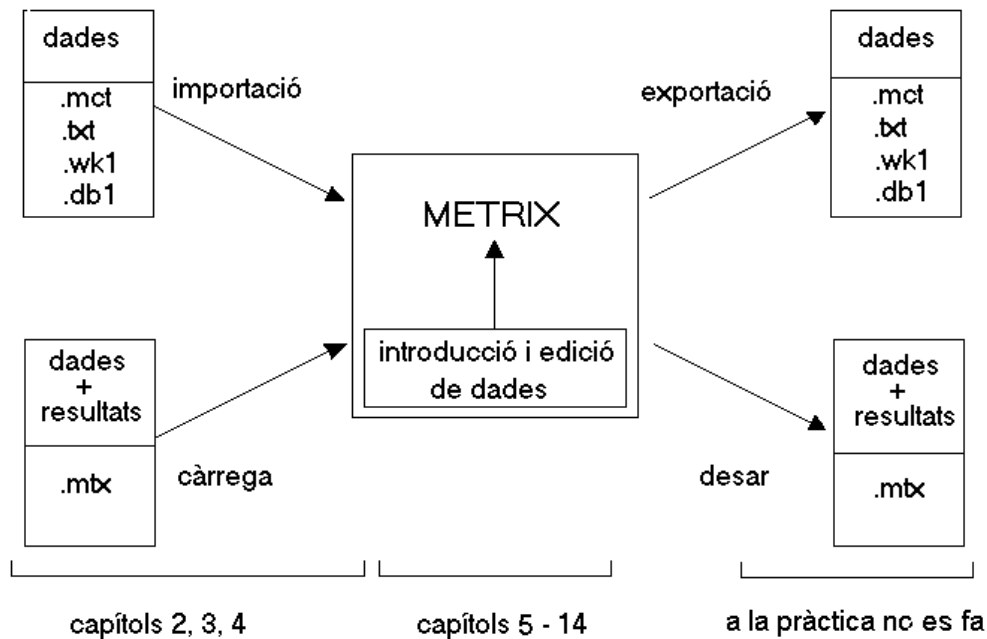
Capítol XIII. Anàlisi d'una prova llarga que pot considerar-se gairebé com un banc d'ítems. Aquesta sessió ens portarà a revisar nous indicadors de l'omissió de l'arxiu *banc.mct*. També fa referència a com enregistrar les puntuacions dels examinands per tal de comparar-les mitjançant les diverses anàlisis i modificacions de la prova original.

Capítol XIV. Aquesta sessió és diferent de les altres, ja que el model de treball no és el convencional. De fet, l'objectiu és fer una rutina d'anàlisi des de la teoria de resposta d'ítem (IRT) i el model de Rasch. Les dades de l'arxiu *tri.mct* serviran per treballar aquest nou enfocament.

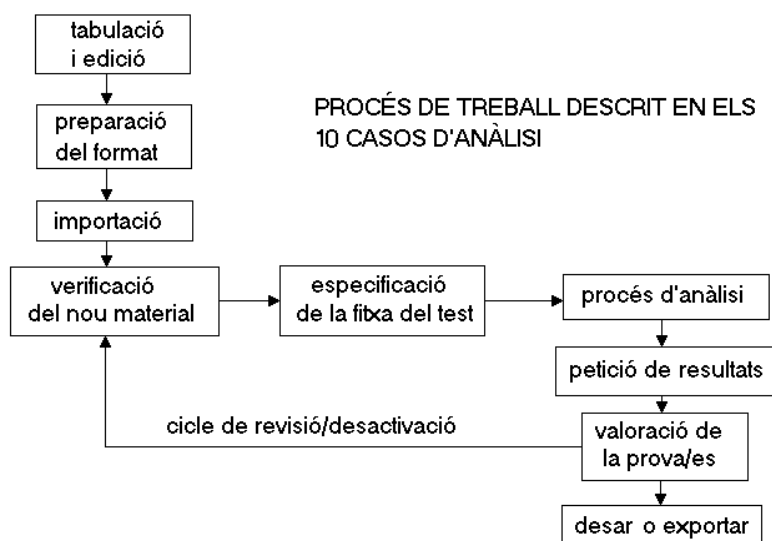
3. RUTINES D'ANÀLISI

Cada capítol presenta un itinerari de treball pràctic. No cal emprar l'ordinador en tots els capítols; de fet, n'hi ha deu de fonamentals que tracten pròpiament de l'anàlisi. Aquests capítols s'han dissenyat sobre una rutina d'anàlisi comuna que ara descriurem. Al marge de les diferències formals i d'objectius, si coneixeu aquesta rutina per sobre podreu situar qualsevol element d'una anàlisi en el procés global de la sessió. Les sessions de treball s'han ajustat aproximadament a una durada de noranta minuts i equips de dos o tres persones assessorades per un grup de monitors de pràctiques. Aquest temps podrà variar segons la dedicació a la revisió de conceptes i l'existència d'errors en els exercicis. En tots els casos, però, és important que vegeu la sessió com una globalitat. Si us manca temps, intenteu acabar la feina en un punt clar que permeti recuperar el fil del procés durant la propera sessió. No es tracta de córrer i sí d'entendre els objectius de cada anàlisi.

Els esquemes següents presenten la rutina bàsica de les sessions de dues maneres complementàries. El primer gràfic situa el treball d'anàlisi (bloc central) entre la importació (càrrega) de fitxers de dades i l'exportació (enregistrament) de les anàlisis (blocs laterals). Qualsevol anàlisi comença amb la captura d'unes dades preparades d'una determinada manera o en un determinat format. METRIX analitza aquestes dades i després permet enregistrar els resultats i els estudis efectuats. Encara que durant aquestes pràctiques poques vegades desareu la feina feta en el disquet, cal saber que el procés normal seria aquest. No enregistrareu les vostres anàlisis, ja que ràpidament s'esgotaria l'espai lliure del disquet. Tampoc empreu el disc dur de l'ordinador, ja que això obligaria a treballar sempre amb la mateixa màquina o bé anar copiant arxius constantment. Així doncs, la rutina bàsica (1) importar/carregar, (2) analitzar i (3) exportar/desar serà aplicada només en part.



Quant al segon esquema, es tracta de veure el caràcter cíclic de l'anàlisi de tests. La captura de dades i l'enregistrament de resultats continuen situats en els extrems d'aquesta seqüència. De fet, esteu veient una ampliació del que en el primer esquema és el bloc central. Qualsevol anàlisi feta amb METRIX demana la verificació inicial de les dades importades i l'especificació d'una sèrie d'instruccions sobre com volem fer l'anàlisi. Després d'haver analitzat les dades, cal avaluar els indicadors de qualitat de la prova. Si tot ha anat bé, s'ha acabat la feina i es poden enregistrar els resultats. Si hi ha cap anomalia, es torna a revisar el material importat i la configuració de l'anàlisi demanada. Això vol dir que després de modificar la configuració de l'anàlisi es repeteix sistemàticament el càlcul de tots els indicadors i resultats en un procés cíclic que va depurant la prova.



Normalment, amb dades reals, poques vegades calen més de tres cicles o revisions. Recordeu que quan parlem de resultats, no ens referim a les puntuacions dels examinands, l'objectiu de les pràctiques és posar a prova les proves. Els resultats seran doncs, d'ara endavant, taules de dades, gràfics i coeficients que ens informen del funcionament intern d'un instrument psicomètric.

Si resumim tot el procés, veiem que (1) METRIX rep unes dades, (2) s'especifica un tipus d'anàlisi, (3) s'avaluen els resultats, (4) es decideix si cal repetir la seqüència d'anàlisi. Totes les sessions tracten aquest procés insistint més o menys en les diverses fases.

4. CONCEPTES

4.1. Anàlisi dels ítems

L'anàlisi dels ítems consisteix a determinar quantitativament la idoneïtat dels ítems que pertanyen a una mateixa prova, considerant cada ítem com la unitat bàsica d'informació del test. L'avaluació del tret psicològic o l'habilitat global mesurada pel conjunt d'ítems dependrà del fet que totes aquestes unitats comparteixin tal objectiu.

Des d'un plantejament convencional, l'anàlisi dels ítems es basa en el càlcul d'una sèrie d'indicadors estadístics sobre les respostes proporcionades per un grup d'examinands. Els més comuns són l'índex de dificultat, el de discriminació i el de dispersió. Cada pregunta o ítem té els seus propis indicadors, que s'hauran de comparar entre si i també amb un valor criteri per tal de valorar la seva idoneïtat. És important recordar que durant la primera fase de qualsevol anàlisi l'objecte d'estudi i atenció no seran els examinands, sinó els ítems i les relacions internes de les respostes que reben.

El tractament estadístic de les dades utilitza les respostes dels examinands (mostres pilot) per revisar la qualitat dels ítems. Més endavant passarà el contrari, ja que amb el material triat podrem avaluar les respostes de cada examinand. Com veieu, es tracta d'un procés cíclic en el qual progressivament construeixim l'instrument de mesura emprant mostres d'assaig fins a arribar a la configuració òptima i final d'ítems.

Malgrat que els diversos indicadors ens assenyalen els ítems amb anomalies, és fonamental contrastar aquesta informació amb la lectura i la interpretació del text de cadascun dels ítems afectats. La decisió sobre la continuïtat o no d'un ítem, o possibles modificacions del seu contingut, depenen d'aquesta comparació. Per a cada sessió o cas seguim la mateixa rutina de treball, i emprem diversos cicles d'anàlisi fins a trobar la configuració idònia de resultats.

4.2. Indicadors i qualitats del test

Índex de dificultat. Consisteix en la proporció de respostes correctes que rep l'ítem. Com a norma, s'ha de procurar que la dificultat sigui centrada, és a dir, al voltant del 0.5 i s'ha d'evitar incloure ítems molt difícils o fàcils. Tanmateix hi ha moltes excepcions a aquesta regla.

Índex de dispersió. Afecta la variabilitat de les respostes i puntuacions que rep cada ítem.

Índex de discriminació. Per detectar el poder separador dels ítems, hi ha molts procediments. N'hi ha, però, quatre d'habituals basats en la correlació ítem-total i el creuament de: (a) dos tipus de coeficient de correlació i (b) dues possibilitats en el càlcul de la puntuació total:

- (a) *coeficient biserial puntual i coeficient biserial,*
- (b) *sense excloure la resposta de l'ítem analitzat del total i exclouent-la.*

Els valors crítics mínims per acceptar els ítems oscil·len al voltant de 0.3, per al coeficient biserial puntual, i de 0.4, per al coeficient biserial, amb qualsevol de les altres dues possibilitats (b).

Puntuació ponderada. És una aplicació de l'índex de discriminació. Consisteix a calcular la puntuació de cada examinand considerant el pes o valor de l'encert aconseguit en cada ítem segons el grau amb què aquest discrimini. Dues persones amb vint encerts tindran la mateixa puntuació directa però diferent puntuació ponderada si no han encertat els mateixos ítems. L'única excepció es produirà quan tots els ítems presentin igual discriminació.

Alternatives incorrectes. Es tracta de comprovar l'homogeneïtat de la distribució de les eleccions rebudes per les alternatives incorrectes de cada ítem. Al marge de si la pregunta és fàcil o difícil, les alternatives incorrectes han d'oferir un atractiu semblant a qui no sàpiga identificar la correcta. Aquesta condició és fonamental, ja que si no es compleix voldrà dir que, encara que l'ítem presenti físicament cinc alternatives de resposta, algunes poden ser tan evidents que ningú les triï. En realitat és com si només n'hi haguessin dues o tres, cosa que fa automàticament més fàcil la decisió d'escollir la correcta. Una anàlisi complementària consisteix a simular què passaria si cadascuna de les alternatives incorrectes fos la correcta. Evidentment el resultat hauria de ser absurd, però sovint no és així, i es descobreixen interferències amb els resultats esperats per a l'alternativa realment correcta. Malgrat que aquest procediment és molt exhaustiu, permet interpretar anomalies impossibles de detectar mitjançant altres indicadors.

Perfil d'encerts. És un gràfic que representa la proporció global d'encerts que rep l'ítem en diferents grups d'examinands amb puntuació total (PT) semblant. A mesura que la PT va augmentant, s'espera que la proporció d'encerts creixi o es mantingui estable, però mai que davalli. De fet, és una manera de confirmar la discriminació de l'ítem sense reduir la informació a un sol valor numèric.

Cada ítem té el seu perfil d'encerts particular o característic que s'ha de revisar per veure que funciona com s'espera. Cal dir que per a l'anàlisi dels qüestionaris només s'apliquen les mínimes possibilitats del perfil d'encerts, ja que aquest tipus de representacions són la base de l'anomenada *corba característica d'ítem*, element fonamental en l'actual anàlisi de tests des de la teoria de resposta d'ítem.

Perfil de les omissions. És un gràfic semblant al d'encerts que representa la proporció global d'omissions que rep l'ítem en diferents grups d'examinands amb PT semblant. Si tractem proves d'encert-error, a mesura que la PT augmenta, s'espera que la proporció decreixi o es mantingui estable, però mai que augmenti. No cal dir que, malgrat això, el cas desitjat per a qualsevol ítem serà la manca de perfil i, per tant, la inexistència d'omissions. Un ítem amb un perfil estable ens diu que, per alguna raó, a l'hora de contestar és difícil d'interpretar i sempre hi haurà qui, al marge de la seva capacitat, el deixa en blanc.

Anàlisi des del model d'escalograma. Informa de la coherència de la pauta de respostes dels examinands, dels ítems i del conjunt del test. L'anàlisi consisteix en la doble ordenació dels ítems (columnes) i examinands (files) segons els encerts obtinguts. S'espera que qualsevol examinand, amb més capacitat o menys, encerti cert nombre de respostes fins a arribar a un punt que, per grau de dificultat, comença a fallar-les. Si hi ha un ajustament al model de l'escalograma, les diverses puntuacions dels qüestionaris només es poden obtenir mitjançant l'encert d'una pauta coherent d'ítems. Dit d'una altra manera, una puntuació baixa ho és per l'encert dels ítems més fàcils i els errors en el més difícils, i no al

revés. De fet, les condicions establertes per aquest model són molt estrictes i mai s'acabaran de complir. Per tant, és una qüestió de grau i així s'han d'interpretar els indicadors anomenats de reproductibilitat (R) dels diversos elements de l'escalograma: els examinands, els ítems i la combinació d'ambdós conjunts.

Fiabilitat de les puntuacions. Considerant ja la correcta adequació dels ítems, és important conèixer també l'error de mesura associat a les puntuacions dels examinands. Tot instrument de mesura té cert grau d'error, és una dada fonamental a l'hora d'interpretar assenyadament els resultats d'un examinand. Realment, mai no tindrem mesures perfectes, ja que sempre hi ha un interval d'error que ens fa treballar amb marges, o franges, delimitats per dues puntuacions extremes. L'amplada d'aquesta franja és la *precisió* de l'instrument, encara que tradicionalment la psicometria ha emprat *fiabilitat* com a terme per avaluar aquesta qualitat dels tests.

La fiabilitat d'un test és un valor relatiu per diversos motius. Es calcula mitjançant una sèrie de coeficients que proporcionen els resultats de manera diferent, encara que sobre una base comuna. Tot coeficient de fiabilitat és una proporció que expressa la relació entre la variància veritable i l'obtinguda en les puntuacions d'un grup d'examinands. Com més s'apropi el coeficient a la unitat millor funciona el test i més petit és l'error estàndard de mesura (SEM o EEM o Se). Rarament, però, trobem valors tan alts, encara que històricament són acceptats, segons diversos autors, coeficients per sobre de 0.75.

El caràcter relatiu del coeficient de fiabilitat també es deu a la mostra emprada. Un mateix test administrat a mostres de poblacions diferents proporcionarà de segur coeficients de fiabilitat també diferents. És important matisar que la fiabilitat, més que una qualitat del test, ho és de les puntuacions assolides per un grup determinat en el test. Cal recordar aquestes condicions i evitar fer generalitzacions més enllà dels resultats obtinguts amb les dades de la mostra emprada.

Per calcular el coeficient de fiabilitat hi ha diversos procediments basats en l'ús del temps, ja que tots pretenen el mateix: comprovar l'estabilitat de les puntuacions administrant el test en dues o més ocasions. La condició bàsica és que el tret o capacitat mesurats no variï durant un temps. Aquest fet és poc habitual en l'avaluació educativa.

Altres procediments només necessiten aplicar el test una vegada. D'una banda hi ha els coeficients que demanen dues o més formes paral·leles del test. Es tracta de construir variants del mateix material, a l'estil de tests bessons, que s'espera que proporcionin per a cada examinand un resultat molt semblant o equivalent. Aquests coeficients s'anomenen d'*equivalència*, ja que ambdues proves, formes o versions, són equivalents i ens confirmen la puntuació obtinguda per a cada examinand. Una variant d'aquest tipus de coeficients és partir el test en dues meitats, considerades equivalents, que imiten el cas de les formes paral·leles. Si hi ha homogeneïtat en els ítems, i la divisió assegura la representativitat del contingut, llavors cadascuna de les meitats o seccions proporciona, per a un mateix examinand, puntuacions semblants a la de l'altra meitat.

La correlació entre les puntuacions d'ambdues meitats, corregida amb la fórmula clàssica de Spearman-Brown, proporciona una estimació del coeficient d'equivalència que probablement hauríem aconseguit si haguéssim construït, des d'un començament, dues formes paral·leles amb la llargada original.

Aquest és el procediment anomenat *dues-meitats* o *split-half*, habitual en la majoria de test (normativa APA), encara que hi ha certes excepcions.

Evidentment, aquest tipus de coeficient dependrà de diversos factors. El més delicat serà el criteri aplicat en fer la divisió del test original, ja que, tot i ser rigorosos, podem crear dues meitats no equivalents i acabar per trobar que el test no és fiable. Per evitar aquest problema, seria recomanable repetir la divisió mitjançant diversos criteris i calcular el coeficient de fiabilitat per observar entre quins valors oscil·la. Fer-ho seria un procés molt llarg, ja que hi ha moltes maneres de partir el test. Una solució alternativa és l'anomenat *coeficient de consistència interna*. Aquest coeficient ofereix una estimació del coeficient que trobaríem en el cas de generar totes les possibles divisions del test i calcular-ne la mitjana.

Tradicionalment, per a ítems puntuats 0-1 (encert-error), es fan servir els coeficients KR-20 i KR-21, tot i que aquest darrer demana més condicions a les dades.

A part d'interpretar el coeficient de fiabilitat, l'usuari d'un test pot calcular i emprar els valors EEM per establir els intervals de confiança en què oscil·len les puntuacions veritables (veritable nombre d'encerts o de punts) d'un grup d'examinands. El procediment és senzill: per exemple, si un examinand ha obtingut vint punts en una prova, i EEM en val tres, caldrà, per treballar només amb un risc d'error del 5%, afegir i restar dues vegades EEM al seu resultat. Fent el càlcul, aquest examinand tindrà la seva *veritable puntuació* entre 14 i 26 punts. Considerarem que el resultat real obtingut és un cas més dels que podria haver assolit dins d'aquest interval. Dit d'una altra manera, en el cas teòric de contestar cent vegades el test només en cinc ocasions, aquest resultat sobrepassaria els límits que abans hem calculat. Evidentment, l'interval de confiança serà més petit (90%, 80%..., 60%) si admetem un grau més gran de risc. No ha de sorprendre el grau d'imprecisió d'un test. Tot i tenir un bon coeficient de fiabilitat, això passarà en quasi tots, ja que en la precisió intervenen molts elements.

Validesa de les puntuacions. La validesa d'un test expressa el grau en què l'instrument mesura el que pretenem. També podem considerar aquesta qualitat com el grau d'utilitat de la prova per detectar allò que ens interessa avaluar en els examinands. Psicomètricament, la validesa presenta tres nivells o enfocaments: la validesa de contingut, de criteri i de constructe.

La *validesa de contingut* no té un coeficient o indicador com el de fiabilitat. Es tracta d'un esforç global en la construcció dels ítems seguint unes pautes establertes. Normalment, però, si una prova supera l'anàlisi d'ítems, i és fiable, el contingut dels seus ítems queda en certa manera garantit. De fet, per exemple, els ítems d'una prova de raonament espacial no creiem que mesurin memòria, capacitat de càlcul, ni habilitats lingüístiques. L'estructura dels ítems i la resposta que demanen són prou evidents per situar-los en el context del raonament espacial.

Malgrat aquesta evidència, durant el procés de construcció de proves psicològiques és habitual comparar els resultats obtinguts en el test pels examinands amb altres puntuacions assolides en proves similars, ja existents i de validesa reconeguda. Aquest és l'enfocament de la *validesa de criteri*, ja que s'utilitzen altres test o dades externes per verificar la validesa del test en creació. Els criteris poden variar i no sempre han de ser puntuacions d'un altre test. El judici d'un expert, la valoració d'un responsable o cap, el diagnòstic d'un psicòleg, són exemples de criteris habituals per validar les puntuacions d'un nou test.

Barems. La interpretació de les puntuacions dels examinands serà relativa. Per a cada individu, caldrà transformar la puntuació directa obtinguda, sumant les respostes, en una altra que informi de la seva situació en relació amb les altres dels membres del seu grup de referència.

El barems o puntuacions equivalents a les puntuacions directes poden ser de molts tipus. Si considerem que la distribució dels resultats per un grup s'ajusta a la corba normal, podrem transformar cada puntuació directe en una nova puntuació típica (z) i aprofitar els avantatges d'aquest model. Si no trobem aquest ajustament, caldrà plantejar una altra estratègia d'elaboració de barems.

Tradicionalment les puntuacions típiques més habituals han estat les anomenades *derivades*, com l'escala T, o bé els centils. Malgrat això, aquestes puntuacions ofereixen una impressió d'extrema precisió

que entra en contradicció amb la franja d'imprecisió d'un test (EEM). Una alternativa habitual és emprar les *puntuacions de classe* que agrupen les derivades en blocs o sectors de la corba normal. El *eneatipus* o *estanines* i els *decatipus* són exemples habituals d'aquesta modalitat de puntuacions, especialment per a tests amb un rang de puntuació inferior a cent punts.

Un altre tipus d'interpretació és considerar la puntuació d'un examinand des d'un enfocament criterial, és a dir, al marge del grup de referència. En aquest cas l'usuari ha de convertir la puntuació d'un examinand en un percentatge segons el nombre màxim d'encerts (ítems) del test. A partir d'aquesta informació, i assumint el grau d'error que inclou, pot comparar aquesta dada amb altres dades que procedixin d'aplicacions següents del test i veure la tendència de l'alumne dins d'una mateixa prova o entre proves diferents. Les proves dissenyades com a tests criterials consideren el nivell estàndard d'aptitud o domini entre el 70% i el 80% d'encerts.

5. RESUM DEL CAPÍTOL

Fins ací la descripció de procediment bàsic de treball durant les sessions. De fet, serà un procés rutinari on les alteracions només dependran del tipus de dades importades o de l'anàlisi que demaneu. Per tal de veureu-ho millor hem exagerat una mica les relacions internes entre les dades, les instruccions i els resultats. Recordeu que aquest cicle de treball no és vàlid únicament per a aquestes pràctiques. Amb noves dades i material vostre podreu fer anàlisis de qualitat que posin en evidència les possibilitats de les proves amb què trebal·leu.

Pel que fa a l'apartat 4, haureu comprovat que hi ha molts conceptes i elements que es tracten progressivament en les sessions de teoria. Tots aniran sortint gradualment i no us han de preocupar ara com ara.