Educar, 17 (1990) 161-181

BIAIX D'ITEMS I HABILITAT EN FRACCIONS

per Joaquim Giménez
Dpt. Educació i Psicologia. Divisió VII. Universitat de Barcelona

RESUMEN

En este artículo se contemplan algunos conceptos clave en la Teoría de respuestas de Items (TRI): curvas características, significado de habilidad y discriminación en dicha Teoría. Se aplica el estudio de niveles de habilidad cognitiva en el aprendizaje de las fracciones en la Educación Básica en 5° y 8° de EGB en España siguiendo el planteamiento de Onslow y Kieren. Los diseños gráficos muestran claramente las diferencias entre edades y cursos.

ABSTRACT

We observe the main concepts about Item Response Theory (IRT): Item characteristic curves, the new meaning of discrimination and ability curves. We apply this ideas to the study of cognitive ability levels with the learning of fractions in elementary school in Spain. The designs showing better than the classic systems the differences between courses (5-8 EGB) and some other ineresting characteristics about the Onslow-Kieren's ideas.

1. Introducció

La recerca actual utilitza encara molts elements de mesura, per obtener informació de les característiques dels subjectes. L'anàlisi clàssica dels instruments de mesura i avaluació ha contemplat que la Maria dominava millor que l'Elisabet l'equivalència de fraccions, ja que a un test determinat responia a un numero més gran d'items favorablement en una prova de les anomenades «objectives». Però, això ens diu que la Maria coneix i domina millor l'equivalència de fraccions? L'habilitat de l'Elisabet és més gran?

Potser que la Maria hagi respost a items poc discriminatius. Es a dir, pot passar que hagi respost a ítems que tant responien els més hàbils i els menys hàbils, i ja no diguem si l'exercici en qüestió ha estat posat a correcuita!

Si l'Elisabet veiem que ha respost a uns quants items més discriminants, potser podrem pensar en primera opinió que ha estat la sort, però ben segur que sabrem més de la seva capacitat real que si tan sols havia respost als items no discriminants. Com saber més sobre la capacitat real que es desprén d'un test?

L'objectiu d'aquestes ratlles és fer una pinzellada del que representa una de les aportacions més importants de la Teoria de Respostes d'items (T.R.I.) a dit problema: generalitzar la idea d'habilitat, i discriminació en un test.

2. De la línia clàssica a un planteig més acurat

A l'esquema clàssic, es parla de caracteristiques dels items, i s'han utilitzat indexos diversos per intentar solucionar aquests problemes. Així: l'index de dificultat (com a proporció d'individus que responen a un determinat item), probabilitat de resposta al atzar, etc. L'ús de correlacions biserials, biserials puntuals, etc. en són exemples clars. La idea que es perseguia era que aquests paràmetres fossin característiques dels items.

Lord (1980) posa en dubte tant des del punt de vista del càlcul com del concepte, la validesa basoluta d'aquests mètodes clàssics. Imainem que a l'item 13, en un grup d'alumnes forts de la classe apareix un index de dificultat 0,85, i al mateix item en un grupet feble s'obté un index 0,36. Està clar que no és una característica pròpia de l'item. Si un grup és molt divers, l'index de dificultat clàssic no és representatiu (Lord-Novick 1968). Les estadístiques són útils únicament en una selecció d'items, quan es construeixen tests per poblacions molt similars a la mostra que es féu servir per obtenir la fiabilitat dels items esmentats.

Anàlogament, la correcció item-test és cada cop més gran si el grup és heterogeni respecte allò que estem mesurant. En canvi, si el grup és més homogeni, pot passar que la correlació sigui petita!

Què vol dir això? Doncs que l'index de correlació no pot ésser considerat com caracteristic de l'item. D'aci que no sigui gaire correcte fer-ho servir com a index que doni el poder discriminatiu d'un item determinat.

De fet, la teoria de generabilitat, millora els conceptes de fidelitat de la teoria clàssica (SPEARMAN 1907) i el paper additiu dels puntatges; així, es parla de: consistència interna, estabilitat, equivalència (com a falta de consistència segons la forma del test, CRONBACH 1972), etc. Però, malgrat disposar de coeficients de variança respecte l'observada, o variança mitjana, o altres, no se solucionen problemes com els provinents de les diferències de puntuacions entre diversos correctors i altres abans esmentats. Pot veure's una reflexió més acurada a HAMBLETON (1985, pg 15-16) o el resum que fa BERTRAND (198).

Aquestes, i altres dificultats intenta salvar-les (o millor, generalitzar-les) la teoria de respostes d'items (Latent trait models*).

3. Biaix d'items

Per HAMBLETON i VAN DER LINDEN (1982) la dificultat en els items i la discriminación, depén de les mostres. La mitjana de nivell d'habilitat i el puntatge d'una mostra influeix de vegades substancialment sobre el valor de l'estadística dels items. D'aci el biaix. De fet, aquests autors consideren cinc aspectes pels quals cal superar el planteig clàssic:

a. La reliability està relacionada directament a la variabilitat de les puntuacions.

b. Els tests realitzats són vàlids per la franja intermitjà, però no pels grups d'habilitat extrema (alta o baixa).

c. Les mesures paral·leles en que es basa la definició de reliability no sempre es compleixen. Es a dir, els individus dificilment es comporten de la mateixa manera en dues passades de tests. Apareixen ansietats i altres factors diferents a la primera vegada, etc (HAMBLETON 1982).

d. No dona una base suficient per determinar com un individu pot reaccionar davant d'un item peculiar del test. Important si es vol dissenyar i predir els puntatges característics a una o més poblacions.

* Qualsevol teoria de la mesura suposa que en situacions de test, les respostes poden ser explicades o predites definint certes característiques (tres = *trait* segons Lord-Novick 1968). Com els trets no són directament mesurables, es consideren els trets latents o habilitats.

e. Classicament, es suposa que la variança dels errors és la mateixa per tots els examinats. No és estrany observar que alguns nois solucionen de vegades algunes tasques de forma més consistent que altres.

Resumint, no es dona una resposta als biaix d'items sobre el màxim poder de discriminació en una escala de puntuacions com tampoc en la conversió de puntuacions d'un test a un altre *(EQUATING OF TEST SCORES)*.

Dels models homogenis a les funcions caracteristiques d'habilitat

Cap als anys 30, es deia que la relació entre habilitat i resposta a un item determinat es describia mitjançant un model de corba normal. Amb posterioritat el model matemàtic s'ha complexificat, de tal forma que s'ha mostrat aquella hipòtesi com a molt particular.

El model actual àmpliament acceptat en T.R.I. es basa en una funció logística, que —certament en alguns casos és molt pròxima a la normal—. Es a dir, representa un model que generalitza la situació anterior i explica amb més claredat els fenómens.

En efecte, un model, especifica una relació entre els trets observables d'una prova i les habilitats no observades vinculades a la resolució del test. Això s'ha fet sempre per mitjà de corbes característiques. D'ací en sorgeixen els models de respostes d'items unidimensionals basats en les dicotomies clàssiques del si i el no (Wright 1968) en els quals és possible estimar l'habilitat amb una mateixa escala per qualsevol subconjunt d'items del domini del model. Aquest model seria homogeni, en el sentit que mesura una habilitat simple. Els avantatges dels mètodes de la Teoria de respostes d'items (TRI) són bàsicament les següents:

a) Si suposem l'existència d'un llarg conjunt d'items que mesuren el mateix dret, l'estimació d'habilitat és independent de la mostra d'items escollida.

b) Si suposem que tractem amb una població ampla, els descriptors (dificultat i discriminació) són independents de la mostra d'individus utilitzada per calibrar els items.

c) S'indica el pes d'habilitat en cada subjecte.

Les corbes característiques, donen la probabilitat de respondre, i aquesta no depén del número de subjectes. Aquesta varia segons el nivell d'habilitat. El més interessant és que aquesta INVARIANCIA no es única a la Teoria TRI.

Aquests mètodes han estat utilitzades darrerament de forma contundent en l'avaluació de l'aprenentatge de la matemàtica, i més en concret a les proves NAEP (National Assessment of Educational Progress 1985).

Fa molt poc que ha estat assumit com a instrument de mesura en l'educació (CARROLL 1988), i basa el seu funcionament en la hipòtesi que la probabilitat d'encertar un item no depén més que del nivell d'habilitat del subjecte i les característiques de l'item. (CARROLL 1988). CARROLL i WILCOX (1982) mostren la utilitat d'aquestes tècniques per definir habilitats com a funcions dels items o també per determinar els distractors en items de respostes múltiples.

La dificultat d'aplicació d'aquests recursos matemàtics i les seves aplicacions informàtiques, fa que requereixen per la seva aplicació óptima d'un nombre mínim de 200 proves, tot i que idealment el número hauria de ser superior al miler (ADAMS 1988, pg 401), en tests de moltes preguntes. L'ús de la Teoria de respostes d'items als grups normals de pocs alumnes és quasi impracticable (CARROLL 1988 citant a Drasgow i Levine).

Les respostes als items (ADAMS 1988) es descriuen —com ja hem dit— per mitjà de corbes (CCI=corbes característiques de l'item) que relacionen la probabilitat del subjecte de donar una resposta correcta $p_i(\theta)$ en funció d'un conjunt de variables:

habilitat del subjecte (θ),
index de probabilitat d'encertar al atzar l'item (c_i),
index del poder discriminatiu de l'item i (a_i),
index de dificultat de l'item i (b_i)

La fòrmula que relaciona aquests paràmetres és (LORD-STOCKING 1988, pg 269)

$$p(\theta)=c_i+\frac{(1-o_i)}{1+e^{-Da_i(\theta-b_i)}}$$

on D és un coeficient constant que usualment val 1,702 i els altres paràmetres varien en els intervals $2\leq a\leq 2$, $0\leq b\leq 0{,}4$ i $0\leq c\leq 0{,}4$.

Podem observar al gràfic següent el tipus de corba i el significat dels diversos paràmetres (BERTRAND, R. 1989).

No tots els items requereixen dels tres paràmetres per descriure'ls. Es parla de models a dos dos paràmetres, quan $c_i=0$. Si, a més a més és constant, és el model de Rasch (LORD-*STOCKING* 1988, pg. 269). Observem que si $c=0$ apareix aleshores la funció normal acumulativa. Aquest paràmetre designa la probabilitat que un alumne de baix nivell respongui correctament a l'item i. Es a dir, l'index de donar una resposta al atzar i encertar-la. Lògicament, un item amb un coeficient c molt baix cal considerar-lo super-fàcil, i és més difícil quan més pròxim estigui d'1. Aquest paràmetre és important en els tests de preguntes múltiples. Més concretament: un item del tipus «cert-

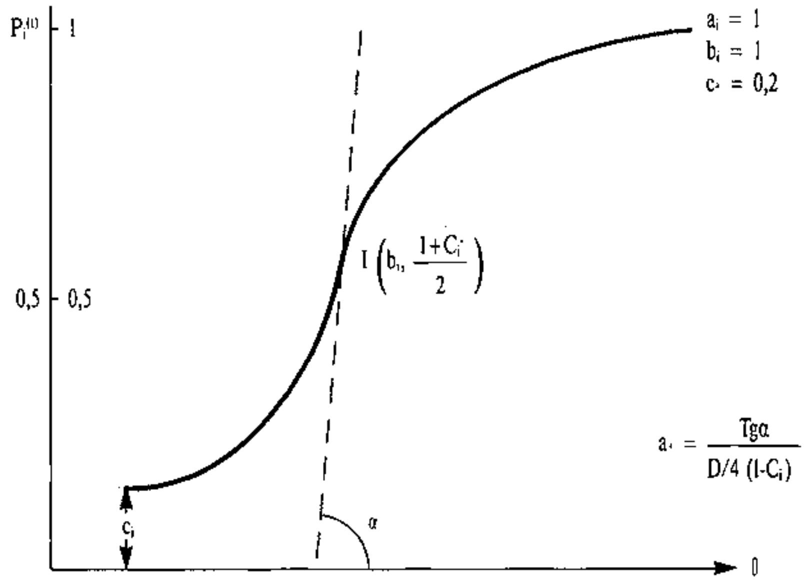


Figura 1. Significat dels paràmetres de la Corba característica d'un item CCI (BERTRAND 1989)

fals», donarà un index d'aproximadament un 50 %. Matemàticament és l'assimptota de la corba a l'origen.

El paràmetre a_i indica el pendent de la corba al seu punt d'inflexió i representa la discriminació de l'item. En tant sigui un pendent alt, els subjectes de baix nivell i els d'alt nivell queden discriminats clarament en la seva puntuació per l'item. Els items de baixa discriminació poden portar a la seva exclussió, segons els nostres objectius.

I el b_i és el nivell d'habilitat al punt d'inflexió. Això indica l'index de dificultat.

La corba corresponent, amb forma del signe de l'integral, va ésser observada ja per BINET-SIMON (1916) entre l'edat i la probabilitat de resoldre un item.

En efecte, la monotonia creixent que pot observar-se, ha estat la indicació clau de la psicologia cognitiva. HULLIN et al. (1983) donen un altre exemple, referint-se a l'habilitat verbal. Diversos autors utilitzen ara aquests mètodes com a model per afinar més sobre els items a considerar en diverses proves i habilitats.

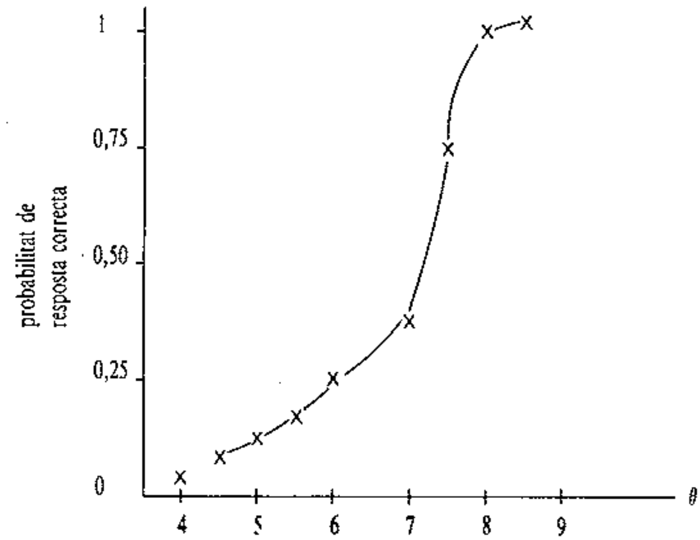


Figura 2. Funció de Binet-Simon, citada per Hambleton-Swaminathan 1985, pg 6.

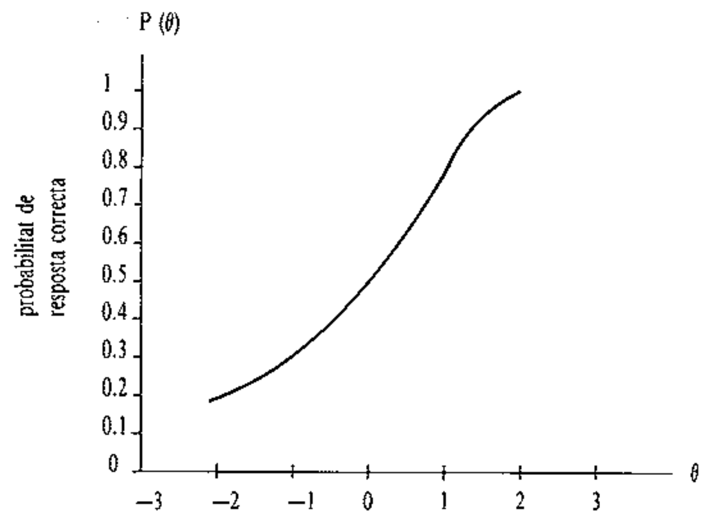


Figura 3. Probabilitat d'encertar un item, en funció de l'habilitat verbal ( ) $a_i = 1{,}07$, $b_I = 0{,}15$

4. Aplicació de T.R.I. a un estudi sobre el coneixement de fraccions

Nosaltres hem aplicat aquests coneixements a la validació i millor comprensió de les dades obtingudes d'un Test de raonament cognitiu (versió KIEREN-GIMÉNEZ 1988) sobre l'adquisició progressiva de la noció de racional positiu.

L'anàlisi factorial dels seus items ens ha facilitat l'observació dels subtests que havien estat introduits per valorar els diversos subconstructes implicats, amb una gran bondat i fiabilitat (GIMÉNEZ 1989, ampliant el treball de RAHIM-KIEREN 1988). Ara, per tal d'exemplificar alguna situació d'aplicació del que hem exposat, mostrarem algunes dades que facin palés el significat dels paràmetres anteriorment citats i el seu ús a la recerca en didàctica de les matemàtiques.

Entre les moltes conclusions del treball més ampli (GIMÉNEZ 1989), comentarem les corbes associades a diversos items, en quant la seva importància per constar el diferent nivell deficultat, diferent poder discriminatiu i probabilitat de resposta al atzar.

Fou emprada una emprada de 959 alumnes de 8è i 945 de 5è d'EGB la província de Barcelona el curs 87-88, significativa de la població per cursos, per estrats de situació i tipus d'escola. Per tal d'actuar sobre un nombre tan elevat de dades, hem utilitzat el programa BILOG (Maximum likelihood item analysis and test scoring logitstic model, Scientifc Software inc. 1983 a l'Uni-

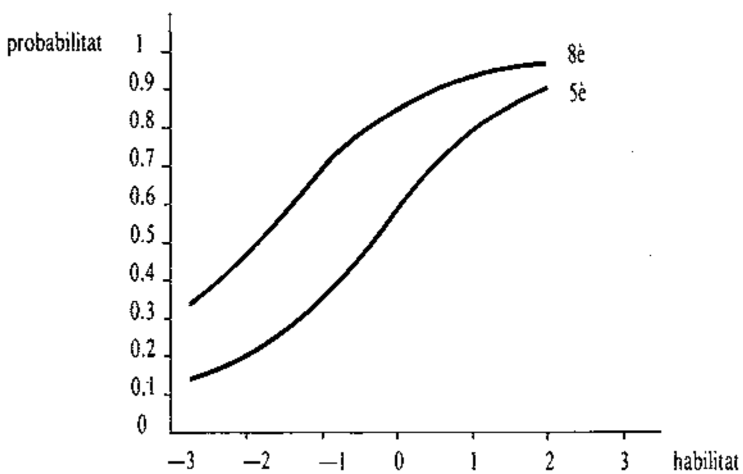


Figura 4. CCI a 2 paràmetres corresponent a la probabilitat d'encert en relació a l'habilitat manifestada en fraccions en el total del test per l'Item 1.

versité Laval, Québec) com un dels recursos informàtics disponibles per anàlisi T.R.I.

Observem l'item 1 del test citat. Repartir dos pizzes entre dues persones. O bé repartir quatre pizzes entre 4 persones. Qui té més. Per què?

Comparen dos cursos: 5è i 8è d'EGB.

La situació d'equivalència —tot i ésser immediata i situada a un primer nivell cognitiu (NOELTING 1982, 1985, BRINDLEY 1981) —es mostra clarament més difícil a 5è que a 8è d'EGB. Si bé els alumnes de 8è nivell d'habilitat —0,5 tenen ja probabilitat 0,95 (quasi seguretat) d'encertar 1 item, amb la mateixa habilitat, a 5è les probabilitats són molt menors. També s'observa que si bé a 8è la probabilitat d'encertar al atzar coincideix amb el 33 % (hi ha tres possibles respostes, de fet), no és el cas a 5è, en que les probabilitats baixen fins el 20 %, i es mostra que no és tan immediat pels nois de nivell baix.

Comparació d'items

Observarem aquesta comparació amb una mateixa pregunta on es vol manifestar l'equilibració (inspirada en NOELTING 1979, 1982, de forma gràfica): «Digues i raona quina combinació xocolata-llet (x,11) té més gust de xocolata?». L'Item 4 compara (7, 4) amb (4, 2) i l'item 6 (5, 3) amb (3, 2). Tenim en compte qui ha donat una argumentació correcta, i no sols haver encertat el criteri de comparació (3 possibilitats).

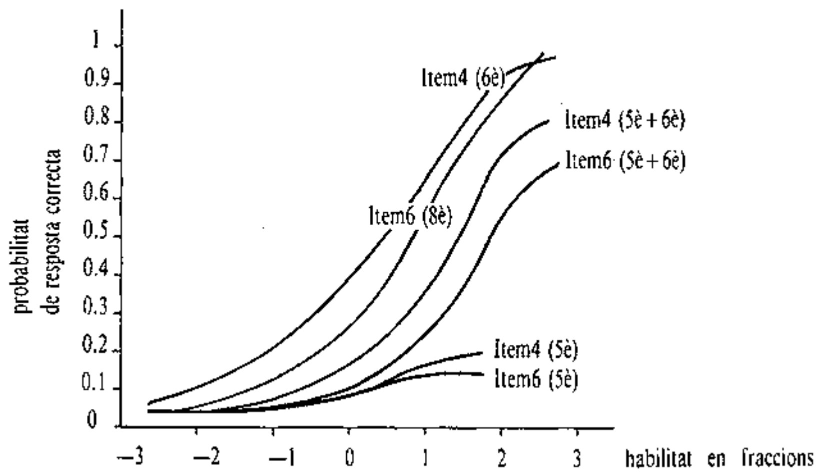


Figura 5. Comparació de CCI a un test de nivell cognitiu de fraccions en dos items i dos cursos diferents per observar el biaix corresponent. (GIMÉNEZ 1989).

L'item 6 té un index de dificultat b_6 més elevat que l'item 4 (comparem ací dos items que l'anàlisi factorial ha unit en un mateix factor: la comparació en proporcions estàtiques com a barreges), amb index de dificultat b_4. La imatge és eloqüent per observar aquest fenomen.

De fet, en molts casos, aquesta observació confirma els percentatges que podiem constatar en un mètode clàssic, però en un mateix gràfic observem també altres característiques. Per exemple, podem constatar la diferencia entre les respostes a 5è i 8è d'EGB pel mateix item. Lògicament apaeix un biaix esperat, ja que —mantenint una major dificultat per l'item 6— aquest és més difícil pels alumnes de 5è que pels de 8è.

El resultat global dels dos cursos, dona també l'ordre de dificultat, i es situa (també de forma lògica) en una situació intermèdia. Tanmateix, els alumnes de nivell d'habilitat alt, el és igual de difícil l'item 4 que el 6 a 8è d'EGB, mentre que a 5è es manté el biaix diferenciador.

No sembla haver gaire diferència des del punt de vista discriminatiu.

Observació de diferent poder discriminatiu

Ho constatarem amb la comparació de dos items aparentment del mateix tipus, on només canvien també les fraccions a comparar. L'objectiu era en el test, constatar l'adquisició de comparacions en la noció de fracció associada a allargaments (operador amb representació continua).

Així, a l'exercici 25 es compara una màquina que transforma una mida de 4 a 1, amb una que redueix de 3 a 1. A l'exercici 26, es demana comparar l'augment de 6 a 10 amb l'augment de 3 a 5.

Observem les corbes característiques TRI corresponents a cada item a la mateixa gràfica.

Hom descobreix al primer item, un pendent aproximat 0,5 i al segon item un pendent 0,9. Es a dir, el primer item és poc discriminatiu, mentre l'altre discrimina amb efectivitat els nois que han descobert l'equivalència dels que no. Es a dir, la comparació entre 1/4 i 1/3 és menys significativa de les dificultats que tenen els alumnes amb aquest tipus de problemes.

Notem, com això no té res a veure amb la dificultat de cada item. Tant és així, que pels nois de nivell d'habilitat alt, el primer item tenen probabilitat més baixa d'encert que l'altre. Al grup de nivell baix, l'item 25 els resulta més fàcil d'assolir.

A partir d'aquests exemples hom pot fer anàlisis dels resultats on els individus de la mostra posseixen un coeficient d'habilitat, i els diversos items es troben associats als coeficients esmentats.

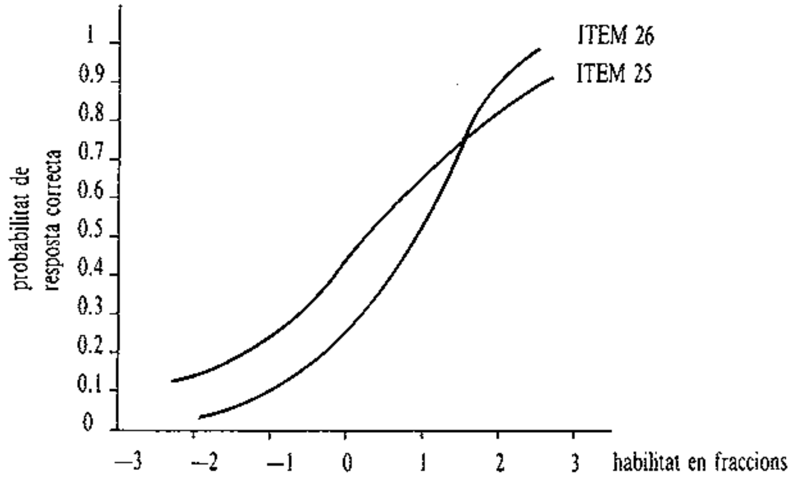


Figura 6. Comparació de CCI en dos items similars per veure el biaix degut a diferent discriminació.

Als gràfics següents, hom pot veure el que ofereix l'ordinador en un dels items que hem comentat, i un tros de la taula que el programa d'ordinador ofereix per tal d'observar-ne els coeficients.

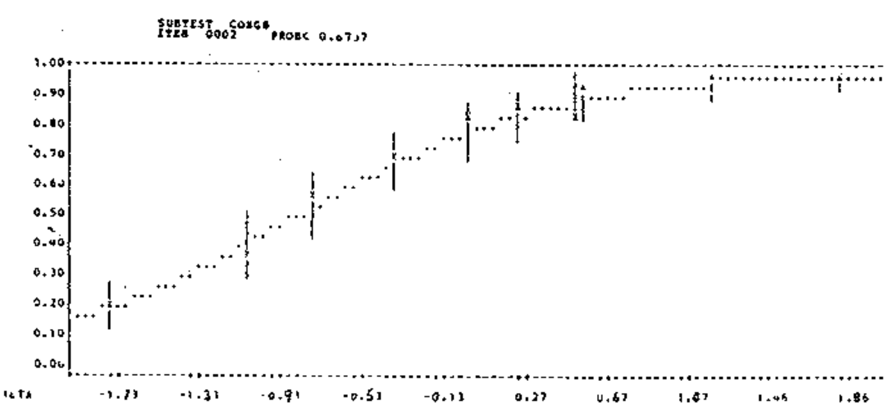


Figura 7. Corba obtinguda per l'ordinador a la primera pregunta del test de nivell cognitiu de fraccions (Giménez 1989).

Item	#Tried	#Right	Pct	Log Ī/1.7	Item-Test Pearson	Correlation Biserial
1	945	886	0,938	1,59	0,281	0,553
2	945	667	0,706	0,54	0,479	0,633
3	945	549	0,581	0,10	0,494	0,624
4	945	259	0,274	—0,57	0,475	0,662
5	945	202	0,214	—0,77	0,518	0,730
6	945	181	0,192	—0,85	0,599	0,728
7	945	916	0,969	2,03	0,256	0,630
8	945	857	0,907	1,34	0,667	0,640
9	945	721	0,763	0,69	0,437	0,602
10	945	558	0,590	0,22	0,460	0,582
11	945	662	0,701	0,50	0,499	0,539
12	945	512	0,542	0,10	0,532	0,668
13	945	393	0,416	—0,20	0,554	0,700
14	945	345	0,365	—0,33	0,543	0,695
15	945	808	0,555	1,04	0,369	0,601
16	945	525	0,556	0,13	0,548	0,690
17	945	527	0,558	0,14	0,513	0,645
18	945	789	0,835	0,95	0,894	0,590
19	945	571	0,604	0,25	0,518	0,404
20	945	199	0,211	—0,78	0,230	0,334
21	945	473	0,501	0,00	0,371	0,464
22	945	376	0,398	—0,24	0,293	0,359
23	945	276	0,292	—0,52	0,377	0,499
24	945	323	0,342	—0,39	0,287	0,371
25	945	449	0,475	—0,08	0,415	0,521
26	945	354	0,375	—0,30	0,557	0,711
27	945	257	0,272	—0,58	0,493	0,661
28	945	134	0,142	—1,06	0,450	0,699
29	945	206	0,218	—0,75	0,503	0,705
30	945	128	0,135	—1,09	0,402	0,632
31	945	239	0,253	—0,64	0,374	0,506
32	945	297	0,314	—0,45	0,367	0,480
33	945	223	0,236	—0,69	0,376	0,515

Taule 1. Indexos donats per BILOG a un test cognitiu de fraccions.

N. preg.	n.i.	n.s.	niv. dific.%	discriminación
1	9	20	72.5	0.55
2	4	20	60	0.8
3	2	18	50	0.8
4	0	10	25	0.5
5	0	6	15	0.3
6	0	5	12.5	0.2
7	12	20	80	0.4
8	6	20	65	0.7
9	1	20	52.5	0.95
10	1	16	42.5	0.8
11	3	19	55	0.8
12	0	14	35	0.7
13	1	12	32.5	0.55
14	0	10	25	0.5
15	2	20	55	0.9
16	0	20	50	1
17	0	14	35	0.7
18	3	20	57.5	0.85
19	3	18	52.5	0.75
20	0	7	17.5	0.35
21	1	14	37.5	0.65
22	0	12	30	0.6
23	0	14	35	0.7
24	0	14	35	0.7
25	0	14	35	0.7
26	0	13	32.5	0.65
27	0	7	17.5	0.35
28	0	13	32.5	0.65
29	0	13	32.5	0.65
30	0	9	22.5	0.45

Taule 2. Dificultat i discriminació dels items de forma clàssica (GIMÉNEZ 1989).

5. Sobre l'habilitat dels nois/es

A la taula següent (resumida de HULLIN et al. 1983, p. 80) podem observar algunes dades que mostren clarament la diferència que pot haver entre els valors clàssics i els obtinguts per T.R.I., que —a la fi— no seleccionaria els mateixos items.

Item	dif. clàss.	dif. TRI	discr. class.	discr. TRI
1	0,89	—1,59	0,71	1,21
2	0,81	—1,19	0,68	1,07
3	0,76	—0,96	0,66	1,09
4	0,82	—1,21	0,66	1,13
5	0,78	—1,11	0,65	0,94
6	0,77	—1,14	0,64	0,82
7	0,76	—1,06	0,64	0,85
8	0,76	—1,09	0,61	0,84
9	0,85	—1,62	0,61	0,85
10	0,91	—2,03	0,60	0,89

Taula 3. Diferenciació entre resultats TRI i clàssics, donat per HULLIN 1983, pg. 1

Després de tot, hom pot concloure que la informació que disposem, ofereix una ponderació a cada item, segons el seu poder discriminatiu (a_i) per a un càlcul més aproximat a la realitat de l'habilitat del subjecte, en test de tipus dicotòmic ($n_i = 1$ encert, $n_i = 0$ error).

La diferència entre la puntuació clàssica i la que s'obté amb aquests factors de correcció es pot observar a les fòrmules següents:

Puntuació clàssica

$$S = \sum n_i \begin{cases} n_i = 1 \text{ si s'encerta} \\ n_i = 0 \text{ si hi ha error} \end{cases}$$

Puntuació a TRI

$$S = \sum a_i n_i \begin{cases} a_i = \text{habilitat a l'item} \\ n_i = \text{index d'encert} \end{cases}$$

Al treball del model de Rasch es diu (HAMBLETON 1982) $b = \Sigma p_{ij} n_i + \varepsilon$ on n_i són les operacions cognitives que influencien la dificultat i p_{ij} són els pasos que reflecteixen la importancia de l'item ε és un factor d'escala.

Es evident, que un alumne pot canviar la seva puntuació, d'acord amb aquestes característiques, però cal dir que rarament seran molt diferents dels resultats clàssics. Ara bé, el model TRI és evidentment més estricte i ajustat a la realitat que plantejavem en iniciar aquest article.

Funció de versemblança i estimació d'habilitat

Com hem dit, la funció matemàtica característica de cada item és funció de l'habilitat. Ara ens proposarem explicar com podem fer una estimació d'aquesta habilitat.

En efecte, si anomenen $e=(e_i)$ al vector de valoració binomial de respostes (1 encert, 0 error), la probablitat de resposta, segons una habilitat, és

$$P(e_i|\theta)=P(e_i=1|\theta)^{e_i}\ P(e_i=0|\theta)^{1-e_i}=P^{e_i}Q^{1-e_i}$$
$$\text{on } Q=1-P_i$$

Si l'esplai és complet, la independència local, dona per cada habilitat θ amb respostes als n items independents,

$$P(e_1......e_n|\theta)=\prod_{i=1}^{n}P(e_i|\theta)$$

A partir de la fòrmula anterior, hom troba fàcilment una relació, que proporciona una funció L de l'habilitat en funció de les probabilits P i Q.

$$L(e_1.....e_n|\theta)=\prod_{i=1}^{n}P^{e_i}Q^{1-e_i}$$

A aquesta funció de θ, s'anomena funció de versemblança.

A partir d'aquesta equació, hom pot prendre logaritmes, i s'obté l'equació següent,

$$\ln L(e|\theta)=\sum_{i=1}^{n}\ln P^{e_i}Q^{1-e_i} \text{ o bé}$$

$$\ln L(e|\theta)=\sum_{i=1}^{n}\left[e_i\ln P_i+(1-e_i)\ln Q_i\right]$$

A la gràfica següent podem observar la funció logaritme de versemblança en els tres casos de TRI (un, dos o tres paràmetres).

El problema és ara estrictament matemàtic. L'equació anteriorment citada compleix l'equació

$$d/d\theta \ \ln L(e|\theta)=0$$

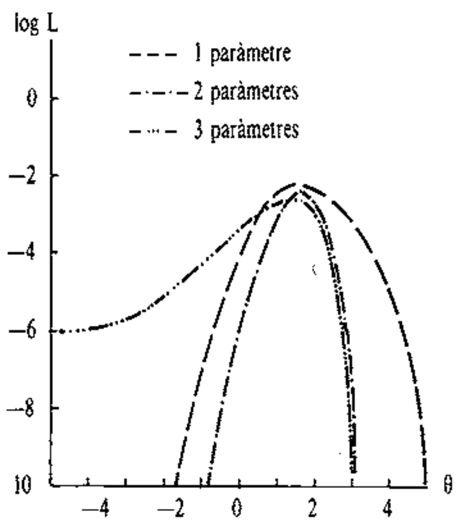


Figura 8. Diferents corbes que relacionen el logaritme de la funció de versemblança amb l'habilitat en els tres casos possibles de paràmetres.

I per tant, el mètode de Newton-Raphson ofereix una solució de la m-ésima aproximació de l'estimació de versemblança, que satisfà l'equació (Hambleton 1985, pag 81) següent

$$\theta_{n+1} = \theta_n - \frac{\left[\dfrac{d}{d\theta}\, l_n\, L(e|\theta)\right]_n}{\left[\dfrac{d^2}{d\theta^2}\, l_n\, L(e|\theta)\right]_n}$$

El procés que es repeteix ofereix una convergència que dona un estimador de l'habilitat. El valor d'estimació, a la m+1 iteracció té un error ε_m que ve donat per la fòrmula (Hambleton 1985, pg 83)

$$\varepsilon_m = \frac{D\left[r - \sum_{i=1}^{n} P_i(\theta_m)\right]}{-D^2\left[\sum_{i=1}^{n} P_i(\theta_m) Q_i(\theta_m)\right]}$$

Evidentment, el càlcul efectiu es fa per mitjà de programes d'ordinador com el que s'ha citat.

Sobre l'habilitat en fraccions

Analitzem ara alguns resultats concrets de la nostra recerca sobre l'habilitat que es deriva del test esmentat. El conjunt de 33 qüestions, ofereix segons la perspectiva clàssica un resultat que podem significar en la gràfica següent sobre l'index d'encert a cada un dels items.

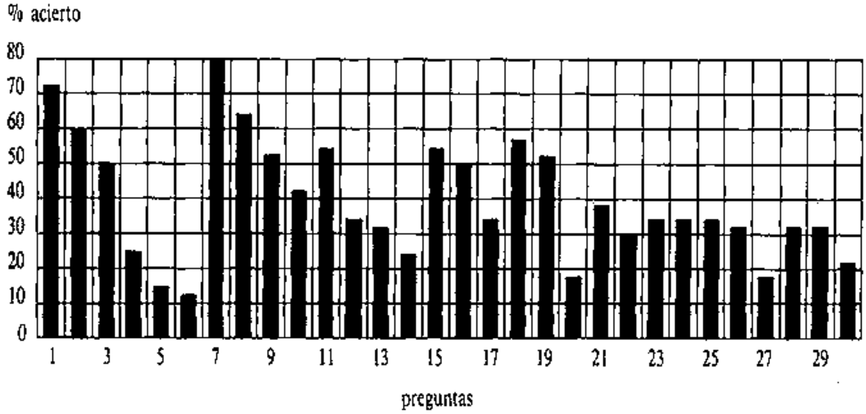


Figura 9. Histograma corresponent als indexos de dificultat clàssics a la prova cognitiva realitzada.

Su.	Resultats (30 items que apareixen)	correct	Percent.	habilitat	Error
A	000000000000000000000000000000	0	0	—3	0,5
B	100000100100001101100001000000	8	24 %	—1,25	0,29
C	111000111100000100000000000000	8	24 %	—0,93	0,44
D	111000111010001111100101000000	14	42 %	—0,37	0,25
E	101000111110001011100101100000	14	42 %	—0,42	0,19
F	100000111111001111011111111010	21	63 %	0,45	0,13
G	111100110101111000101111011100	21	63 %	0,48	0,18
H	110000111111001111111111111000	22	66 %	0,46	0,14
I	111111111111111111001101100000	22	66 %	1,06	0,41
J	100000111111111111111111111110	26	78 %	1,24	0,27
K	111101111111111111111010111111	27	81 %	1,50	0,35
L	111111110111111111011110011111	28	84 %	1,92	0,44
M	111111110111111111011011111010	28	84 %	1,59	0,40
N	111111111111111111111010111110	30	90 %	2,14	0,38
O	111101111111111111111110111110	30	90 %	1,94	0,44

Taula 4. Alguns resultats significatius de diversos alumnes, percentatge de respostes correctes i índex d'habilitat en fraccions.

A la taula anterior, com a contrast, observem com alumnes que han respost una mateixa quantitat d'items, apareixen amb nivells d'habilitat diferents a l'interval [—3, +3]. S'han escollit uns quants subjectes significatius de les diferències que ofereix la TRI, i els errors corresponents en els coeficients.

El concepte d'habilitat que hom pot observar quantificadament, no és anàleg a l'index de respostes correctes. Què indica, doncs? L'assignació d'un coeficient relacionat amb la dificultat dels items que aquell individu ha respost.

6. Conclusions

Entre les múltiples aplicacions d'aquest projecte, es considera que són importans:

— l'adaptació d'un test a l'habilitat dels subjectes
— la detecció de biaix d'items deguts a diverses variables.

L'exemple que hem ofert mostra la diferència de curs. En un altre moment podem fer intervenir el sexe, nivell cultural, etc.

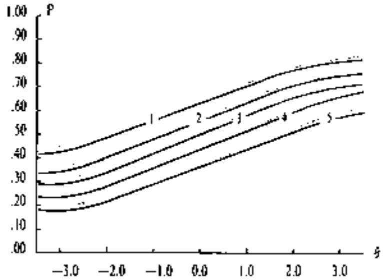
Es possible detectar un biaix, examinant la diferència entre la corba d'un item per una població i la característica de l'item per una subpoblació. La superposició de corbes, és indicadora de l'absència de biaix.

Si les corbes es tallen, vol dir que l'item té un biaix diferenciat a cada nivell d'habilitat θ.

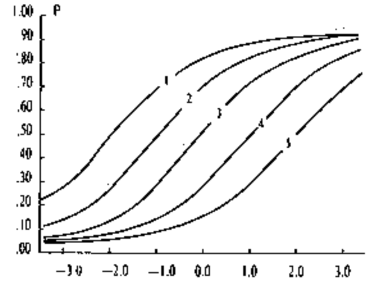
La evidència sobre la imatge gràfica dels diferents conceptes i valors és l'element més satisfactori d'aquesta anàlisi, que possibilita una visualització dels resultats dels test incomparable fins el moment!

Apendix 1

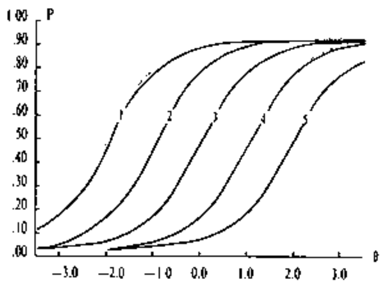
A les gràfiques següents, observarem unes quantes comparacions per veure les formes de les corbes característiques en relació a diversos paràmetres. Això pot ésser il·lustratiu de les interpretacions donades al text.



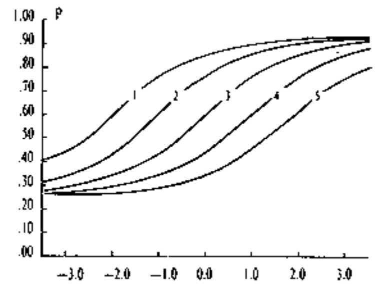
Item	b	a	c
1	−2.00	.19	.00
2	−1.00	.19	.00
3	0.00	.19	.00
4	1.00	.19	.00
5	2.00	.19	.00



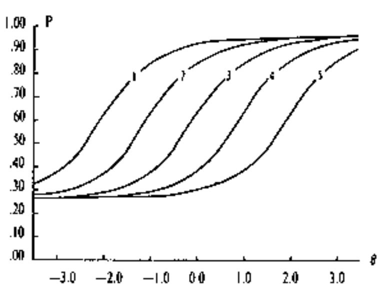
Item	b	a	c
1	−2.00	.39	.25
2	−1.00	.39	.25
3	0.0	.39	.25
4	1.00	.39	.25
5	2.00	.39	.25



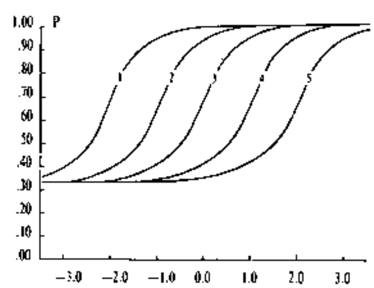
1	−2.00	.59	.00
2	−1.00	.59	.00
3	0.0	.59	.00
4	1.00	.59	.00
5	2.00	.59	.00



1	−2.00	.99	.25
2	−1.00	.99	.25
3	0.0	.99	.25
4	1.00	.99	.25
5	2.00	.99	.25



1	−2.00	.99	.00
2	−1.00	.99	.00
3	0.0	.99	.00
4	1.00	.99	.00
5	2.00	.99	.00



1	−2.00	1.39	.25
2	−1.00	1.39	.25
3	0.0	1.39	.25
4	1.00	1.39	.25
5	2.00	1.39	.25

Apendix 2. Breu sintesi cronològica de la Teoria T.R.I.

1916	Utilització de gràfics en el desenvolupament de tests sobre nivells d'habilitat en relació a una variable independent. BINET i SIMON.
1936	RIRCHARDSON troba relacions entre paràmetres del model TRI i els paràmetres clàssics.
1943-44	LAWLEY ofereix procediments nous per l'estimació de paràmetres.
1952	LORD descriu el model ogival normal a 2 paràmetres.
1957-58	BIRNBAUM substitueix els models logístics pels models normals.
1960	RASCH desenvolupa a Europa la teoria a 3 paràmetres.
1967	WRIGHT popularitza i catalitza la teoria als EE.UU.
1968	Moments forts de la recerca de LORD i NOVICK sobre LATENT TRAIT.
1969	S'utilitza el programa BICAL per eixamplar els mètodes de RASCH.
1974	Programa LOGIST a dos paràmetres.
1977	Molts investigadors comenten els trets importants de TRI a Journal for Educational Measurement.
1980	WEISS reuneix elements fonamentals de l'ús de TRI.
1982	LORD i el seu equip desenvolupen un nou model de programa LOGIST.
1985	Desenvolupament del programa a 3 paràmetres en BILOG.

REFERÈNCIES

ADAMS, R; ROWE, K.J. Item Bias in Keeves, J. (ed) *Educations Research, methodology and measurement.* Pergamon Press. Oxford. 398-403. 1988.

ANDERSEN, J. et al. An evaluation of Rasch's structural model for test items. *British Journal of mathematical and Statistical Psychology.* 78, 31, 34-98. 1968.

BERTRAND, R. Pourquoi de nouvelles théories de la mesure? a *Mesure et évaluation en Education.* vol. II, n° 2. U. Laval. Québec, 1988.

BINET; SIMON. *The development of intelligence in young children.* The training school. Vineland. NJ. 1916.

BIRNBAUM, A. Some latent trait models and their use in inferring an examenee's ability. In F.M. LORD & M. NOVICK (ed). *Statistical theories of mental test scores.* Reading MA Addison Wesley. 1968.

BRINDLEY, J. et al. Report of Calgary Junior High School mathematics Project. Univ. Alberta. 1981.

CARROLL, J.B. Measurement theory. Future development in educational measurement in Keeves, J. (ed). *Educational Research, methodology and measurement.* Pergamon Press. Oxford. 247-269. 1988.

CRONBACH, L.; GLESER, G.; RAJARTNAM, N. *The dependability on behavioral measurements.* John Wiley. N. York, 1972.

GIMÉNEZ, J. *About continuous operator subconstruct in rational number* in G. VERGNAUD (ed) Proceedings PME 15. Paris, 1989.

GIMÉNEZ, J. *Innovación metodológica sobre el racional positivo.* Tesis. 1989.

HAMBLETON, R.; VAN DER LINDEN. Advances in Item response. Theory and applications: An introduction. *Applied Psychological Measurement* 1982, 6, 373—378. 1982.

HAMBLETON, R.; SWAMINATHAN, H. *Item response theory Principles and applications.* Kluwer Nijhoff Publ. Boston Mass. 1985.

HULLIN, C.; DRASGOW, F.; PARSONS, C. *Item response theory Applications to psychological measurement.* Dow -Jones Irwin, Homewood. Illinois, 1983.

LORD, F.M. *Applications of item response theory to practical testing problems.* Lawrence Erlbaum Ass. Hillsdale. N. York, 1980.

LORD, F.; NOVICK. *Statistical theories of mental test scores.* Reading Mass Addison Wesley. 1968.

LORD-STOCKING. Item response theory in Keeves, J. (ed). *Educational research, methodology and measurement.* Pergamon Press. Oxford. 269-272. 1988.

NOELTING, G. The development of proportional reasoning and the ratio concept. Part 1. Differentiation of stages. *Educational Studies in Mathematics* vol. II, p. 331-363. 1980.

NOELTING, G. *Le development cognitif et le mécnisme de l'equilibration.* Gaëtan Morin (ed) Québec. 1982.

NOELTING, G. (dir) et al. *Partage. Cahier clinique. Echelle ordinalle partant sur les notions de rapport et de fractions.* Centre de Psychologie cognitive et developpementale. Univ. Laval. Québec, 1984.

RAHIM & KIEREN, T. A preliminary report on the reliability and factorial validity of the RNTT in the Rep. of Trinidad and Tobago. in Behr, M (ed) *Proceedings of PME-NA 88 Illinois*; 114-120. 1988.

SPEARMAN, C. Demonstration of formulae for true measurement of correlation in *Americal Journal of Psichology* 18, 161-169. 1907.

WILCOX, R. Some empirical and theoretical results on an answer-until-correct scoring procedure. *Br. J. Math. Stat. Psychol.* 35: 57-70. 1982.

WRIGHT, B. Sample free test calibration and person measurement. *Proceedings of the 1967 Invitational Conference on testing problems.* Princenton NJ Educational Testing Service. 1968.