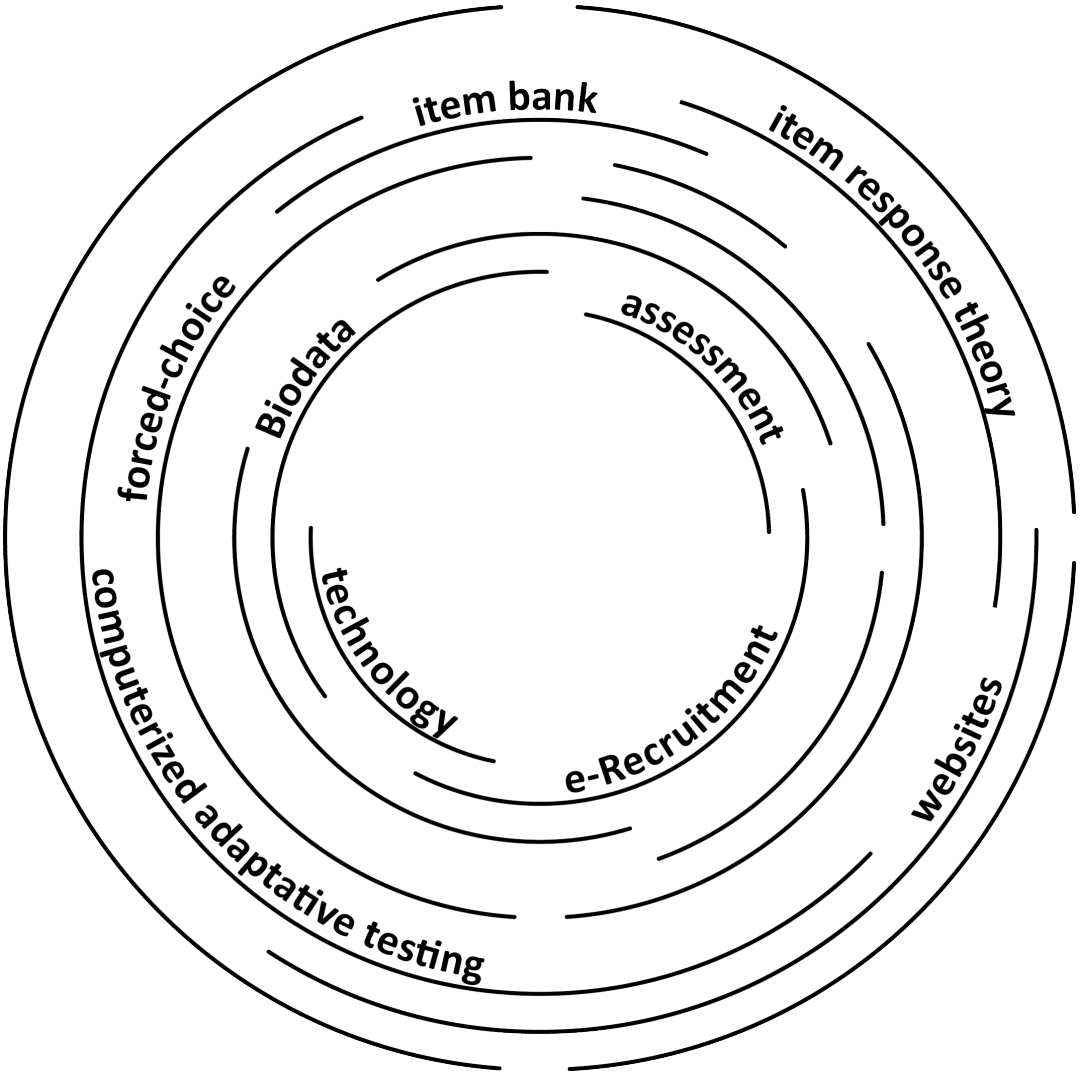
VIII SEMINARIO

AVANCES EN SELECCIÓN DE PERSONAL: NUEVAS TECNOLOGIAS EN EVALUACIÓN Y MEDICIÓN



CÁTEDRA DE
MODELOS Y
APLICACIONES
PSICOMÉTRICOS

UAM-IIC

PROGRAMA

8:45 **Recogida del material del seminario.**

9:15 **Apertura y presentación.**

D. Francisco Santolaya

Presidente del Consejo General de Colegios Oficiales de Psicólogos

D. Alberto Barrientos

Director General del Instituto de Ingeniería del Conocimiento

9:30 **Conferencia.**

Technology and assessment

Dra. Nancy T. Tippins

Vicepresidenta Senior de la compañía Corporate Executive Board (CEB Inc)

10:40 **Sesión de pósters.**

Café.

11:40 **Conferencia.**

Case studies on using technology and assessment

Dra. Nancy T. Tippins

Vicepresidenta Senior de la compañía Corporate Executive Board (CEB Inc)

12:40 **Conferencia.**

El reclutamiento de personal mediante recursos web

Dr. Antonio L. García-Izquierdo

Profesor Titular de la Facultad de Psicología de la Universidad de Oviedo y Director de la Cátedra Universitaria Asturias Prevención

13:10 **Conferencia.**

Avances en medición adaptativa

Dr. Francisco J. Abad

Profesor Titular de la Facultad de Psicología de la Universidad Autónoma de Madrid

13:40 **Debate y clausura del seminario.**

El objetivo de los seminarios de la Cátedra MAP es fomentar y contribuir a la medición psicológica de calidad, especialmente en el campo de las organizaciones.

En este octavo seminario se ha escogido como tema central Avances en selección de personal: Nuevas tecnologías en evaluación y medición.

—

24 de junio de 2016

Facultad de Psicología

Universidad Autónoma de Madrid

TECHNOLOGY AND ASSESSMENT

Dra. Nancy T. Tippins
Vicepresidenta Senior de la compañía
Corporate Executive Board (CEB Inc)

Technology and Assessment

Nancy T. Tippins
CEB

June 24, 2016

CEB
WHAT THE BEST COMPANIES DO

What I tried to say!

Buenos días.

Estoy muy contenta de estar aquí con ustedes en Madrid y hablar de uno de mis temas favoritos – pruebas, evaluación, y tecnología.

Mi español es terrible. Estoy tratando de aprenderlo, y lo poco que se decir son frases como “una cerveza, por favor” que no son muy útiles hoy, por tanto voy a hablar en inglés.

CEB 2

What is technology-enhanced assessment?

- Any assessment in which one or more component is facilitated by technology – administration, scoring, reporting, data storage
- A technology-enhanced assessment can be used in personnel assessment for selection, development, certification, evaluation

CEB 3

What is technology-enhanced assessment?

Different forms of technology use different forms of software to administer and score different types of test that evaluate different kinds of constructs.

Multiple Forms of Technology
- Computers
- Laptops
- Tablets
- Phablets
- Smart Phones
- Telephones
- Video
  - Still pictures
  - Animation
  - Cameras for monitoring
- Audio

Multiple Forms of Software
- Test administration
  - Computer adaptive testing
  - Branching
- Scoring tools
  - Scanners
- Dynamic norms/reports
- Report generators
- Storage and retrieval of data
- In-basket tools
- Assessment portals
- Data mining tools

Multiple Test Formats
- Multiple choice
- Matching
- Constructed response
  - Short answers – sentence completion; background questionnaire
  - Written responses – business case
- No format (i.e., Big Data)

Multiple Constructs
- Single KSAO
- Multiple KSAOs
  - Personality scales
  - Arithmetic word problems
  - Work sample
- Narrowly defined
  - Conscientious-ness
- Broadly defined
  - Problem solving

CEB 4

Overview

Purpose
- Define the critical questions about technology-enhanced assessments
- Present the research that answers those questions
- Discuss the implications for employment testing programs
- Highlight some of the unanswered questions that remain

CEB 5

- Is the assessment valid and reliable?
- Does technology affect adverse impact?
- Does cheating on unproctored internet tests (UIT) matter?
- Does the testing environment matter?
- What impact does a technology-enhanced assessment have on the applicant pool?
- Should we use mobile testing?
- Are scores taken on different devices equivalent?
- How do candidates react to technology-enhanced assessments?
- What are the advantages and disadvantages of technology-enhanced assessments?
- Does UIT present an ethical problem?
- What else can I worry about?

CEB 6

Is the assessment valid and reliable?

CEB

Validity

The validity of high fidelity, technology-enhanced work samples and simulations is almost unquestioned.

- Many of the work samples and simulations are validated using a content-oriented strategy; there are few estimates of criterion-reported validity.
- The meta-analysis of Schmidt & Hunter (1998) estimated the validity of work samples to be 0.54.
- Roth et al. (2005) estimated the meta-analytic validity of work samples to be 0.34.
  - The relatively high validity for high-fidelity simulations suggests that simulations that embed a relevant form of technology would also have high validity.
    - In a local validation study, Cucina et al. (2013) found lower validities for the dimensions of a video-based test against a number of criteria.
      - Uncorrected validities ranged from 0.096 to 0.114 (0.148 to 0.180 corrected) for a training criterion and from 0.104 to 0.166 (0.155 to 0.246 corrected) for a work sample criterion.

CEB 8

Validity

- Lievens & Patterson (2011) found that validities of low-fidelity simulations were similar to those of high-fidelity simulations (0.36 versus 0.37, uncorrected; 0.54 versus 0.56, corrected) when job performance was used as a criterion.
  - The similarity of validities is particularly important from a practical perspective, considering the cost of administering many high-fidelity simulations.
- There is also evidence that higher fidelity in both the stimulus (Lievens & Sackett, 2006) and the response (Lievens et al., 2012) leads to greater validity.

CEB 9

Validity

- Reliability of technology-enhanced assessments.
  - The use of humans to evaluate work samples and simulations requires administrative guidelines to ensure consistency across applicants, assessors, time, location

CEB

10

Validity

- In contrast, the validity of a UIT measuring cognitive ability is frequently questioned.
  - Does a UIT have lower validity than the proctored version of the test?
    - UITs result in test taker cheating and distractions; thus they are believed to have lower reliability and thus lower validity.
    - Research does not always confirm this assumption.
      - Kaminski & Hemingway (2009) found comparable validities for proctored and unproctored tests.
      - Beaty, Nye, Borneman, Kantrowitz, Drasgow, & Grauer (2011) meta-analysis showed the validities of a proctored test and unproctored test to be comparable.
  - Does a UIT have sufficient validity to warrant its use in selection?
    - Most published studies show statistically significant relationships between the test score and the criterion measure.

CEB

11

Questions for Practitioners

- Should we use UIT despite the questions about validity?
- How reliable is a single individual's test score?
- Do the benefits of high fidelity simulations outweigh the costs?

CEB

12

Does technology affect adverse impact?

CEB

Adverse Impact in Employment Testing

- In the U.S., adverse impact refers to the proportion of the minority test takers passing a test relative to the proportion of the majority test takers.
  - Adverse impact is related to group mean differences but it is not the same thing.
- For testing purposes, minorities are defined by race and sex.
- If adverse impacts, employers must be able to demonstrate that their selection procedures are job relevant and "consistent with business necessity."
  - This usually means that that the test can be linked to an essential knowledge, skill, ability, or personal characteristic and predicts some outcome that is important to the business (e.g., performance, turnover).

CEB 14

Adverse Impact

- Because of the extreme sensitivity of such data, adverse impact is rarely reported.
  - Occasionally, researchers compare group mean differences.
    - With the exception of mobile testing, most do not report sizeable differences in *d*s
- Keep in mind:
  - Large sample sizes usually mean statistically significant differences and UIT usually results in larger applicant pools.
    - If the increase in sample size is due to more qualified applicants who are part of the majority group taking the test or more unqualified applicants who are part of the minority group, adverse impact would be exacerbated.

CEB 15

Adverse Impact

- There is some evidence that simulations reduce adverse impact.
  - Olson-Buchanan et al. (1998) and Schmitt & Mills (2001) found that video assessments reduced adverse impact.
  - Cucina et al. (2013) reported no adverse impact in their video-based test measuring interpersonal skills, judgment/decision making, cooperativeness/sensitivity, emotional maturity, and task orientation.

CEB 16

Does cheating on unproctored internet tests (UIT) matter?

CEB

Does Cheating Occur?

- There appears to be a small amount of cheating on unproctored tests.
- The extent of cheating found in the literature may depend on how cheating is defined.
- Hense, Golden, and Burnett (2009)
  - Effect size of .32 between scores on a proctored and unproctored job simulation.
- Arthur, Glaze, Villado, and Taylor (2009)
  - 7.7% of a sample cheated on speeded cognitive ability test (cheating defined as 1 SEM)
- Lievens and Burke (2011)
  - Small d scores on cognitive test scores across four levels of jobs
    - Some in the opposite direction expected (proctored scores were higher than unproctored scores)
    - <2.2% exhibited negative score change (only those in the high end of the distribution were retested)
- Nye, Do, Drasgow, & Fine (2008)
  - No differences in scores from unproctored and proctored versions of an internet version of a perceptual speed test

CEB 18

How Can Cheating Be Detected?

- Direct observation of test taker's behavior
- Changes in means, standard deviations, pass rates
- Unusual response patterns (e.g. test taker gets easy items wrong and hard items correct)
- Statistical detection (e.g., Guo & Drasgow, 2010; Makransky & Glas, 2011; Tendeiro et al., 2013)

CEB 19

Can Cheating Be Prevented?

- Proctoring through technology
  - Evidence suggests that proctoring reduces cheating more on tests with right and wrong answers than on self-report inventories (Viswesvaran & Ones, 1999)
  - A number of authors point out that proctoring is not perfect (Bartram, 2009; Drasgow et al., 2009; Foster, 2009)
  - Cameras, including video records
  - Key strokes
    - Prohibited keys, e.g., print screen
    - Inactivity
    - Rhythm of key strokes to detect substitutes
- Verification testing
- Alternative forms
  - LOFT
  - CAT
- Serial presentation of responses (Foster, 2009)

CEB 20

Can Cheating Be Prevented?

- Various verbal statements have been found effective.
- Warnings
  - Cheating detection mechanisms are in place and there are consequences to cheating
  - Dwight & Donovan, 2003; Fan et al., 2012; Landers et al, 2011 McFarland, 2003
- Reminders that responding honestly is in the best interest of the individual, the selection process is fair, and the test taker is moral
  - Pace & Borman, 2006
- Honesty contracts
  - Burke, 2009

CEB 21

Important Considerations Related to Cheating on a UIT

- Proctored testing often does not meet the gold standard for test administration (Drasgow, Nye, Guo, & Tay, 2009; Bartram, 2009; Foster, 2009)
  - Comparison may be between some cheating and a lot of cheating
- Cheating may not occur at equal rates across the entire distribution of test scores
  - Do smart people need to cheat? (Only people in the high end of the distribution are typically retested.)
- Practice effects in second administration (usually the proctored test) minimize the amount of cheating observed
  - Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard (2007) found effect size of .26 on measures of GMA
- Higher motivation in the proctored setting (one of the later stages of the hiring process) may exist

CEB 22

Important Considerations Related to Cheating on a UIT

- The candidate's physical and psychological health and ability to focus on the test may vary across time.
- Fewer distractions occur in the proctored setting, minimizing the extent of cheating observed
- Better quality computer equipment may be used in the proctored setting
- The real threat of verification testing may serve as a significant, effective deterrent to cheating

CEB 23

Cheating on Other Forms of Tests

- How do people cheat on work samples and simulations?
  - Pre-knowledge of materials
  - Outside assistance
  - Violation of time limits

CEB 24

Questions for Practitioners

- How accurate is any one person's score?
  - How can we tell if a specific person cheated or not?
- Prevent or detect cheating: what is the better approach?
  - What are our best deterrents?
    - CAT
    - LOFT
    - Verification testing
    - Warnings
  - What are our best methods for detection?
    - Proctors
      - Human
      - Electronic
        - Cameras
        - Keystrokes
    - Computer algorithms
    - Self Report
    - ???

CEB 25

Questions for Practitioners

- How bad is it to hire someone who cheats on a test?
  - How much of a performance decrement will actually occur?
  - Is cheating on an employment test related to other dishonest behavior?
    - If I cheat on an employment test will I falsify company records?
- Is cheating on a cognitive ability test worse than distorting responses on a personality inventory?
- Is the "cost" of cheating worth the "benefit" of expanding the applicant pool?
- What can be done (or should be done) to prevent cheating in proctored settings?

CEB 26

Does the testing environment matter?

CEB

Testing Environment

- Why does the test environment matter?
  - Accuracy of the individual's test score
  - Validity and reliability of the test
  - Applicant reactions
  - Size of the applicant pool
- What effect does the testing environment have on test taker performance?
  - Weiner & Morrison (2009)
    - Testing environment is rated least favorably when test is administered in unproctored, off-site compared to on-site proctored and unproctored.
    - Non-cognitive scores are related to ratings of test environment.

CEB 28

Professional Guidance on Testing Environment

Standards for Educational and Psychological Testing

Standard 5.4

The testing environment should furnish reasonable comfort with minimal distractions.

Comment: Noise, disruption in the testing area, extremes of temperature, poor lighting, inadequate work space, illegible materials, and so forth are among the conditions that should be avoided in testing situations. The testing site should be readily accessible. Testing sessions should be monitored where appropriate to assist the test taker when a need arises and to maintain proper administrative procedures. In general, the testing conditions should be equivalent to those that prevailed when norms and other interpretative data were obtained.

CEB 29

Professional Guidance on Testing Environment

Standard 5.6

Reasonable efforts should be made to assure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent means.

Comment: In large-scale testing programs where the results may be viewed as having important consequences, efforts to assure score integrity should include, when appropriate and practicable, stipulating requirements for identification, constructing seating charts, assigning test takers to seats, requiring appropriate space between seats, and providing continuous monitoring of the testing process. Test developers should design test materials and procedures to minimize the possibility of cheating. Test administrators should note and report any significant instances of testing irregularity. A local change in the date or time of testing may offer an opportunity for fraud. In general, steps should be taken to minimize the possibility of breaches in test security. In any evaluation of work products (e.g., portfolios) steps should be taken to ensure that the product represents the candidate's own work, and that the amount and kind of assistance provided should be consistent with the intent of the assessment. Ancillary documentation, such as the date when the work was done, may be useful.

CEB 30

Professional Guidance on Testing Environment

Principles for the Validation and Use of Personnel Selection Procedures

Administration Environment - There are a number of factors that potentially affect test administration: appropriate workspace, adequate lighting, and a quiet, comfortable setting, free of distractions. The researcher should consider these conditions and their potential effects on test performance. At a minimum, selection procedure administration should be in an environment that is responsive to candidates' concerns about the selection procedures and maintains their dignity.

CEB

31

Questions for Practitioners

- Are our professional guidelines for test administration aligned with today's UIT practices?
- Who is responsible for ensuring a proper testing environment?
  - Should the employer act *in loco parentis* and dictate the test environment?
  - Should the employer provide the conditions under which most people do their best and allow the candidate to decide what conditions work best?
  - The correct answer is not clear and probably depends on a number of factors:
    - The consequences of cheating and "bad" test scores
    - The need for large applicant pools

CEB

32

What impact does a technology-enhanced assessment have on the applicant pool?

CEB

Applicant Pool

- How does UIT affect the applicant pool in terms of size and diversity?
  - UITs generally increase the size of the applicant pool
    - Applicant testing is not limited by time or location
  - Applicant pool may increase due to undesirable applicants
    - Some applicants are "window shopping"
    - Some lack the commitment to make further efforts related to employment
    - Some applicants are practicing
- Could the pool of capable candidates shrink because some candidates withdraw because of the practice of UIT?
  - Some applicants do not want to work for a company that uses UIT and tolerates assumed cheating
  - Some applicants do not have access

CEB 34

Applicant Pool

- Is there a digital divide?
  - Do some groups have greater access to technology?
- How great is the digital divide?
  - In May, 2013, the Pew Research Center and American Family Life Project found
    - 91% of Americans own a cell phone
    - 34% own a tablet
    - 61% and 58% own laptop or desktop computers
  - 2011 data show
    - Men and women use the internet about the same rate
    - Blacks and Hispanics use the internet somewhat lower than Whites
    - Younger people use the internet more than older people

	% of adults who use the internet August 2011	% of adults who access the internet via broadband at home
All adults (age18+)	78	62
Men	80	65
Women	76	59
Race/ethnicity		
White, Non-Hispanic	80	66
Black, Non-Hispanic	71	49
Hispanic	68	51
Age		
18-29	94	76
30-49	87	70
50-64	74	60
65+	42	30

CEB 35

World Internet Usage

WORLD INTERNET USAGE AND POPULATION STATISTICS
NOVEMBER 30, 2015 - Update

World Regions	Population (2015 Est.)	Population % of World	Internet Users 30 Nov 2015	Penetration (% Population)	Growth 2000-2015	Users % of Table
Africa	1,158,355,663	16.0 %	**330,965,359**	28.6 %	7,231.3%	9.8 %
Asia	4,032,466,882	55.5 %	**1,622,084,293**	40.2 %	1,319.1%	48.2 %
Europe	821,555,904	11.3 %	**604,147,280**	73.5 %	474.9%	18.0 %
Middle East	236,137,235	3.3 %	**123,172,132**	52.2 %	3,649.8%	3.7 %
North America	357,178,284	4.9 %	**313,867,363**	87.9 %	190.4%	9.3 %
Latin America / Caribbean	617,049,712	8.5 %	**344,824,199**	55.9 %	1,808.4%	10.2 %
Oceania / Australia	37,158,563	0.5 %	**27,200,530**	73.2 %	256.9%	0.8 %
WORLD TOTAL	**7,259,902,243**	**100.0 %**	**3,366,261,156**	**46.4 %**	**832.5%**	**100.0 %**

http://www.internetworldstats.com/stats.htm

CEB 36

Internet Users by Country (2016)

#	Country	Internet Users (2016)	Penetration (% of Pop)	Population (2016)	Non-Users (internetless)	Users 1 Year Change (%)	Internet Users 1 Year Change	Population 1 Y Change
1	China	721,434,547	52.2 %	1,382,323,332	660,888,785	2.2 %	15,520,515	0.46 %
2	India	462,124,989	34.8 %	1,326,801,576	864,676,587	30.5 %	108,010,242	1.2 %
3	U.S.	286,942,362	88.5 %	324,118,787	37,176,425	1.1 %	3,229,955	0.73 %
4	Brazil	139,111,185	66.4 %	209,567,920	70,456,735	5.1 %	6,753,879	0.83 %
5	Japan	115,111,595	91.1 %	126,323,715	11,212,120	0.1 %	117,385	-0.2 %
6	Russia	102,258,256	71.3 %	143,439,832	41,181,576	0.3 %	330,067	-0.01 %
7	Nigeria	86,219,965	46.1 %	186,987,563	100,767,598	5 %	4,124,967	2.63 %
8	Germany	71,016,605	88 %	80,682,351	9,665,746	0.6 %	447,557	-0.01 %
9	U.K.	60,273,385	92.6 %	65,111,143	4,837,758	0.9 %	555,411	0.61 %
10	Mexico	58,016,997	45.1 %	128,632,004	70,615,007	2.1 %	1,182,988	1.27 %
11	France	55,860,330	86.4 %	64,668,129	8,807,799	1.4 %	758,852	0.42 %
12	Indonesia	53,236,719	20.4 %	260,581,100	207,344,381	6.5 %	3,232,544	1.17 %
13	Viet Nam	49,063,762	52 %	94,444,200	45,380,438	3.3 %	1,564,346	1.07 %
14	Turkey	46,196,720	58 %	79,622,062	33,425,342	5.1 %	2,242,750	1.22 %
15	Philippines	44,478,808	43.5 %	102,250,133	57,771,325	4.4 %	1,855,574	1.54 %
16	South Korea	43,274,132	85.7 %	50,503,933	7,229,801	1.2 %	522,375	0.42 %
17	Italy	39,211,518	65.6 %	59,801,004	20,589,486	1.7 %	666,922	0.01 %
18	Iran	39,149,103	48.9 %	80,043,146	40,894,043	7.7 %	2,784,831	1.18 %
19	Spain	37,865,104	82.2 %	46,064,604	8,199,500	2.2 %	805,002	-0.12 %
20	Pakistan	34,342,400	17.8 %	192,826,502	158,484,102	9.7 %	3,024,054	2.07 %
21	Canada	32,120,519	88.5 %	36,286,378	4,165,859	1.8 %	559,167	0.96 %
22	Egypt	30,835,256	33 %	93,383,574	62,548,318	3.3 %	990,548	2.05 %
23	Argentina	30,359,855	69.2 %	43,847,277	13,487,422	3.1 %	904,688	0.99 %
24	Thailand	29,078,158	42.7 %	68,146,609	39,068,451	6.2 %	1,708,982	0.28 %
25	South Africa	28,580,290	52 %	54,978,907	26,398,617	3.9 %	1,078,982	0.9 %

http://www.internetlivestats.com/internet-users-by-country/

Internet Usage in Spain

- 37,865,104
- Internet Users in Spain (2016*)
- Share of Spain Population: **82.2 %** (penetration)
- Total Population : **46,064,604**

http://www.internetlivestats.com/internet-users/spain/

CEB

38

Internet Usage in Spain

Year	Internet Users**	Penetration (% of Pop)	Total Population	Non-Users (Internetless)	1Y User Change	1Y User Change	Population Change
2016*	37,865,104	82.2 %	46,064,604	8,199,500	2.2 %	805,002	-0.12 %
2015*	37,060,103	80.4 %	46,121,699	9,061,596	5.1 %	1,814,825	-0.3 %
2014	35,245,278	76.2 %	46,259,716	11,014,438	5.9 %	1,967,122	-0.42 %
2013	33,278,156	71.6 %	46,455,163	13,177,007	2.2 %	720,809	-0.39 %
2012	32,557,347	69.8 %	46,637,082	14,079,735	3.1 %	982,491	-0.15 %
2011	31,574,855	67.6 %	46,708,366	15,133,511	3 %	911,074	0.23 %
2010	30,663,782	65.8 %	46,601,492	15,937,710	6.1 %	1,775,583	0.66 %
2009	28,888,199	62.4 %	46,295,191	17,406,992	5.8 %	1,581,258	1.04 %
2008	27,306,942	59.6 %	45,817,016	18,510,074	9.6 %	2,391,965	1.34 %
2007	24,914,976	55.1 %	45,209,538	20,294,562	11.1 %	2,481,223	1.51 %
2006	22,433,753	50.4 %	44,537,926	22,104,173	6.8 %	1,436,094	1.56 %
2005	20,997,660	47.9 %	43,854,761	22,857,101	10.5 %	1,999,741	1.59 %
2004	18,997,918	44 %	43,167,276	24,169,358	12 %	2,037,545	1.63 %
2003	16,960,373	39.9 %	42,475,265	25,514,892	98.9 %	8,434,196	1.58 %
2002	8,526,178	20.4 %	41,815,486	33,289,308	13.9 %	1,043,365	1.42 %
2001	7,482,812	18.1 %	41,230,518	33,747,706	34.8 %	1,930,668	1.18 %
2000	5,552,144	13.6 %	40,749,800	35,197,656	94 %	2,690,314	0.88 %

* estimate for July 1, 2016
** **Internet User** = individual who can access the Internet at home, via any device type and connection.
Source: ***Internet Live Stats*** (www.InternetLiveStats.com)
Elaboration of data by *International Telecommunication Union (ITU), World Bank, and United Nations Population Division.*

Mobile Devices: Smart Phones and Tablets

- In the U.S., ownership of both types of devices is increasing
  - Blacks and Hispanics own smart phones at a higher rate than whites
  - Rate of ownership of tablets is about the same for all ethnic groups
- Demographic groups most likely to own tablets include:
  - Those living in households earning at least $75,000 per year (56%), compared with lower income brackets
  - Adults ages 35-44 (49%), compared with younger and older adults
  - College graduates (49%), compared with adults with lower levels of education

CEB

40

Mobile Phones

Rankings	Country or regions	Number of mobile phones	Population	Connections/100 citizens
	World	6,880,000,000+	7,012,000,000[illegible]	97
01	China	1,276,660,000[illegible]	1,364,270,000[illegible]	93.2
02	India	1,017,968,757[illegible]	1,295,291,543[illegible]	80.30
03	United States	327,577,529	317,874,628[illegible]	103.1
04	Brazil	284,200,000	201,032,714[illegible]	141.3
05	Russia	256,116,000	142,905,200[illegible]	155.5
06	Indonesia	236,800,000	237,556,363	99.68
07	Nigeria	167,371,945	177,155,754	94.5
08	Pakistan	150,000,000[illegible]	180,854,781[illegible]	83[illegible]
09	Bangladesh	133,720,000[illegible]	157,497,000[illegible]	84.95
10	Japan	121,246,700	127,628,095	95.1
11	Germany	107,000,000	81,882,342	130.1
12	Philippines	106,987,098	94,013,200	113.8
13	Mexico	101,339,000	112,322,757	90.2
14	Iran	96,165,000	73,973,650	130
15	Italy	88,580,000	60,790,400	147.4
16	United Kingdom	83,100,000	64,100,000	129.6
17	Vietnam	72,300,000	90,549,390	79
18	Turkey	72,200,000	81,619,392	92.5
19	France	72,180,000	63,573,842	114.2
20	Egypt	93,670,000	87,120,000	107.17
21	Thailand	69,000,000	67,480,000	105
22	Ukraine	57,505,555	45,579,904	126.0
23	South Korea	56,004,887	50,219,669	111.5
24	Spain	55,740,000	47,265,321	118.0
25	Argentina	56,725,200	40,134,425	141.34

https://en.wikipedia.org/wiki/List_of_countries_by_number_of_mobile_phones_in_use

CEB

41

Applicant Pool

- Does the composition of the applicant pool change with the use of technology-enhanced assessments?
  - Answers are difficult to come by because few comparative studies (technology vs. no technology) are done.
  - There is no reason to believe that the use of technology-enhanced assessment in proctored environments has any effect on the composition of the applicant pool.
  - Although the size of the applicant pool generally increases with UIT, we do not know
    - If the demographic composition changes
    - If the distribution of scores changes
  - The higher rate of ownership by Blacks and Hispanics suggests members of these groups have greater access to testing.
- Limitations of equipment and connectivity may have an impact on who
  - Can access a test
  - Can complete a test

CEB

42

Size of Applicant Pool, Cheating, and Performance

Landers & Sackett (2012)

- Simulation study of effects of cheating and increases in size of applicant pool on performance
- Conclusions:
  - More cheating results in lower mean criterion scores, regardless of other factors.
  - If UIT increases the applicant pool, adopting UIT can result in higher mean criterion performance among those selected even if cheating occurs.

CEB

43

Questions for Practitioners

- Does the use of technology in assessment limit the accessibility for all groups?
  - How widespread is access to technology in the geographic location of interest?
- Does familiarity with technology have any effect?
- How does the use of technology in assessment affect the disabled?

CEB

44

Should we use mobile testing?

CEB

Mobile Testing

- Much has been made over testing on mobile devices, but is it really desired?
- Some HR Professionals believe mobile testing is appropriate (Fallaw & Kantrowitz, 2013)
  - 23% believe candidates want access to application forms and assessments via mobile devices
  - 43% would provide assessments via smartphone if the option existed
  - 41% indicated they would allow candidates to take assessments via smart phone if research showed the comparability of test scores from mobile devices with those on computers
  - 23% indicated that allowing candidates to take assessments on mobile devices was inappropriate

CEB 46

Mobile Testing

- Only a small percent of people take cognitive tests on mobile devices.
  - Hawkes (2013)
    - 1% of applicants take a graduate level cognitive test using a mobile device and 7% use a mobile device for a non-cognitive assessment.
  - Doverspike, Arthur, Taylor, & Carr (2012)
    - 1.7% of over 1M applicants used a mobile device to complete an assessment consisting of measures of verbal and numerical ability and a five-factor model measure
      - Caucasian (53.8%), female (59.0%) and under the age of 40 (74.3%)
      - Test takers using the mobile device performed less well on cognitive measures than those using a PC; however, no differences were found on the non-cognitive measures.
  - Morelli, Illingworth, Scott, & Lance (2012)
    - Also found no differences when the assessment task involved a non-cognitive test.

CEB 47

Implications of Mobile Devices on Assessment

Characteristics of Mobile Device	Implications for Testing	Impact on Candidate
• **Mobile devices are mobile - they can be easily transported**	• Candidate may take test wherever he/she chooses	• Candidate may take test in less than ideal conditions • Cheating may be easier in some environments than others • The applicant pool may expand • Little impact on scores on self-description inventories is observed • Scores on cognitive measures are typically lower when the candidate takes the test on a mobile device
• **Mobile devices require an internet connection to access test content**	• The reliability and speed of the internet connection varies from place to place	• Interrupted internet connections may - reduce the number of candidates completing the test - require procedures for restarting or resuming a test • Variations on speed of the internet connection may - affect scores on speeded tests - frustrate or distract candidates - reduce the equivalency of scores from tests on stationary computers and from mobile computers • Bandwidth may not be sufficient for large downloads such as video-based test content • Charges for data transmissions may be a deterrent to some candidates testing

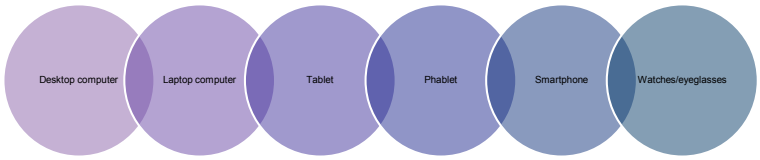
Implications of Mobile Devices on Assessment

Characteristics of Mobile Device	Implications for Testing	Impact on Candidate
• Laptops, tablets, and smart phones typically have smaller screens than desk top computers and or poorer resolution	• Small screens, particularly those on smart phones, may reduce the visibility of some item content • Small screens may also make responding more difficult	• Test scores may be affected by the visibility of the item content due to difficulties in reading material in small fonts and difficulty in manipulating the user interface • A need for excessive scrolling may use up test time or frustrate the test taker • Closely positioned responses options may lead to inaccuracies in responding • Differences in visibility and ease of responding may reduce the equivalency of scores from tests on stationary computers and from mobile computers • Poor screen resolution may slow a test taker down and have a negative impact on test score, particularly when the test is speeded.

CEB

49

Equipment, Internet Connection, and Bandwidth



Why do different devices have different effects on test performance?

- Device Characteristics
  - Screen size
  - Interface
  - Permissibility

→

- Cognitive demands
  - Working memory
  - Information processing
  - Psychomotor ability
  - Distractibility/selective attention

CEB

50

Mobile Devices

- O'Connell, Arthur, & Doverspike (2015) Summary of Mobile Testing Literature
  - 3/11 papers are peer-reviewed
  - Operational data with very large samples
  - <5% use mobile devices; no distinction between smart phones and tablets
  - African Americans, Hispanics, and Women use mobile devices more often
  - Range of test methods: self-report Likert items, biodata. SJT, multiple-choice, multimedia/simulation
  - No concerns about measurement equivalency for both cognitive and non-cognitive constructs

CEB

51

Mobile Devices

- O'Connell, Arthur, & Doverspike (2015) Summary of Mobile Testing Literature (continued)
  - Mean differences between scores from different devices
    - No mean differences on non-cognitive constructs
    - Limited research but substantial mean differences found when studied
  - Group mean differences
    - Paradoxically, mean differences are not larger with mobile devices; seem to be smaller
      - Arthur, Edwards, and Barrett (2002) and Edwards and Arthur (2007) report similar findings for constructed-response tests
      - Mobile devices will likely exacerbate subgroup differences if historically lower performing demographic groups have a disproportionally and substantially higher representation in mobile device usage for assessment

CEB 52

Mobile Devices

- O'Connell, Arthur, & Doverspike (2015) Summary of Mobile Testing Literature (continued)
  - Conclusion: Smartphones produce different scores because they are
    - Most extreme on all the critical characteristics/variables
      - Smallest screen
      - Most difficult/challenging response interface
      - Widest permissibility, including the most unsuitable locations
    - Thus, collectively engenders the highest level of cognitive load/demands
    - Score differences may be reduced by app-based design solutions but unlikely to eliminate differences
      - Differences in cognitive scores may be reduced by app-based designs but not eliminated
      - Mobile devices will probably exacerbate subgroup differences compounded by fact that lower performing demographic groups have higher mobile device usage

CEB 53

Questions for Practitioners

- Should we assess individuals on mobile devices?
  - Do the advantages outweigh the disadvantages?

CEB 54

Are scores taken on different devices equivalent?

CEB

Score Equivalency

SJI

- Potosky & Bobko (2004)
  - Modest degrees of cross-mode equivalence between scores from paper and pencil and internet-based tests
- Ployhart, Weekley, Holtz, & Kemp (2003)
  - Variance-covariance matrices are not equivalent for scores from paper and web-based forms

Biodata

- Ployhart, Weekley, Holtz, & Kemp (2003)
  - Variance-covariance matrices are not equivalent for scores from paper and web-based forms

CEB

56

Score Equivalency

Personality Test

- Ployhart, Weekley, Holtz, & Kemp (2003)
  - Variance-covariance matrices are not equivalent for scores from paper and web-based forms measuring conscientiousness, agreeableness, and emotional stability
- Mead, Michels, and Lautenschlager (2007)
  - Comparability of test scores on some (including conscientiousness) but not all personality constructs.
  - Lack of measurement invariance when those with choice of medium are compared to those without a choice.
- Chuah, Drasgow, & Roberts (2006)
  - Found equivalence between traditional paper and pencil measures of personality and internet measures from unproctored testing sessions
- Salgado, & Moscoso (2003)
  - Found equivalence of scores form paper and pencil measure of a Big Five personality measure to an internet based version

CEB

57

Score Equivalency

Cognitive ability

- Mead and Drasgow (1993)
  - Equivalence between scores from paper-and-pencil and computer administered tests for power cognitive ability test
  - No equivalence between scores from each medium on speeded cognitive ability tests
  - No differences between adaptive and conventional tests administered via computer.
- Potosky & Bobko (2004)
  - Modest degrees of cross-mode equivalence between scores from paper and pencil and internet based cognitive ability tests (timed).

CEB 58

Score Equivalency

- Other thoughts on score equivalence:
  - Many organizations infer equivalency by comparing means and standard deviations, rather than establishing measurement invariance across constructs (Vandenberg & Lance, 2000)
    - Sample size is often a significant problems
    - Comparisons across multiple devices and multiple cultures compound the difficult
  - Few studies compare high fidelity simulation to a paper-and-pencil test.
    - Implicit assumption is that these are measuring something different.
    - Chan & Schmitt (1997)
      - Compared written and video SJI – not an equivalence study; attributed different candidate reactions to the reading comprehension demand in the written form

CEB 59

Questions for Practitioners

- How do we conduct equivalence studies on small samples?
- How do conduct equivalence studies that take into account device as well as culture?

CEB 60

How do candidates react to technology-enhanced assessments?

CEB

Candidate Reactions

- Candidate reactions have long thought to have an impact on behavioral outcomes such as
  - Job offer acceptance (Boudreau & Rynes, 1985; Murphy, 1986)
  - Consumption of the organization's products and services (Rynes and Barber, 1990)
  - Reduction of the applicant pool(Boudreau & Rynes, 1985; Murphy, 1986)
  - Probability of a challenge to the selection procedures (Rynes, 1993; Rynes et al., 1980; Smither et al., 1993).
- Research does not confirm the relationships between applicant reactions and variety of outcomes
  - No relationship between applicant perceptions of the testing process and applicant withdrawal (Ryan et al., 2000; Truxillo et al., 2002; Hausknecht et al.'s, 2004)

CEB 62

Candidate Reactions

- Why?
  - Too few studies
  - Too few outcomes included
  - Global measures of candidate reactions
    - Ryan & Huth (2008) emphasize the need for specificity in evaluating candidate reactions to various components of an assessment
- Sackett and Lievens (2008) characterized this lack of evidence for a relationship between applicant reactions and individual or organizational outcomes as "the Achilles hell of this field" (p. 439).

CEB 63

Candidate Reactions

- Candidate reactions appear to be highly dependent on the type of technology-enhanced assessment:
  - Simulations and work samples generate more favorable reactions to simulations and work samples than to paper-and-pencil tests (Schmidt, Greenthal, Hunter, Berner, & Seaton,1977; Smither et al., 1993; Rynes & Connerly, 1993; Hausknecht, 2004)
  - Only 2/11 of studies on mobile testing reported by O'Connell et al. (2015) reported candidate reactions
    - Gutierrez & Meyer (2013) found mobile testing is perceived as less fair, more difficult, less comfortable; makes individual no more likely to apply; does not indicate employer provides not better place to work; applicants show preference for PC
    - Smeltzer (2013) found mobile testing interfered with opportunity to perform; did not improve perceptions of organization; candidates did not think organization would view taking test on mobile device negatively or affect hiring decision; preference for PC
  - Continuous monitoring via technology is perceived as more invasive and threatening to privacy; applicants may prefer one-time check such as biometric identification (Karim & Kaminsky; 2013)

CEB 64

Questions for Practitioners

- What is the appropriate methodology for evaluating applicant reactions?

CEB 65

What are the advantages and disadvantages of technology-enhanced assessments?

CEB

The advantages and disadvantages of technology-enhanced assessment have been discussed extensively although not necessarily in the published literature.

Advantages
- Costs
  - Test administrators
  - Printing
- Ease of updating
- Location of test – convenience; expansion of the applicant pool
- Applicant reactions – simulations, games
- Candidate engagement – simulations
- Speed in processing
- Image/employment brand
- RJP – high fidelity

Disadvantages
- Costs
  - Software
- Potential for cheating and other forms of malfeasance
- Impact on validity
- Location of test – distractions
- Applicant reactions – CAT
- Privacy concerns
- Candidate engagement – virtual assessments

???
- Costs
  - Equipment
  - Internet connection
  - Bandwidth
- Applicant pool
- Diversity
- ADA accommodations
- Equivalence of scores
- Effect on adverse impact

CEB 67

Costs

- Few people report the costs of their technology-enhanced assessments.
- What are the sources of costs?
  - Development
    - Psychological assistance (e.g., validation research, item development and research)
    - IT Support (programming, maintenance of infrastructure, security activities)
    - Actors and videographers
  - Maintenance
  - Management
  - Administration
  - Equipment and internet connections
  - Scoring
  - Reporting
  - Per applicant fees (increase with the number of applicants)

CEB 68

Benefits

- More people report the benefits of their use of technology in assessment
- Few people report the benefits of their technology-enhanced assessments in precise terms.
  - One exception is Cucina et al. who did not report costs or include estimates of SDy but did share utility estimates. The average utility for a video-based assessment tool (using only significant validities) was $109,183,823.
- What are the sources of benefits?
  - Reduced costs of administration (TAs, scorers, reporters)
  - Increased applicant pool
  - Better applicant pool
  - More efficient/timely staffing
  - More positive candidate reactions to high fidelity simulations (Schmidt, Greenthal, Hunter, Berner, & Seaton,1977; Smither et al., 1993; Rynes & Connerly, 1993; Hausknecht, 2004)
  - Additional data (e.g., response times)
  - Company image

CEB 69

Cost/Benefit

Final thoughts:

- Most organizations don't have a firm idea of the costs or the value of the benefits.
  - Even the tangible costs/benefits are difficult to evaluate; the intangible ones are almost impossible
- The importance of cost in determining what test to use is not clear.
- Different factors have different weights for different organizations.
  - For example, some value the image they present to candidates more than others.

CEB

70

Questions for Practitioners

- How important is cost relative to goals for the assessment program?
  - What are the benefits that offset the costs?
- How much effort should be put on quantifying costs and benefits to technology-enhanced assessments?
- If it's important to make a business case for technology-enhanced assessment, what is the appropriate methodology? What are the most important factors to consider?

CEB

71

Does UIT present an ethical problem?

CEB

Ethics

- What does our Ethics Code require us to do?

(a) Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques.

(b) Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation.

(Section 9.02, Use of Assessments)

CEB 73

Ethics

Psychologists who develop tests and other assessment techniques use appropriate psychometric procedures and current scientific or professional knowledge for test design, standardization, validation, reduction or elimination of bias, and recommendations for use. (Section 9.05, Test Construction)

Psychologists who offer assessment or scoring services to other professionals accurately describe the purpose, norms, validity, reliability, and applications of the procedures and any special qualifications applicable to their use. (Section 9.09a, Test Scoring and Interpretation Services)

Psychologists make reasonable efforts to maintain the integrity and security of test materials and other assessment techniques consistent with law and contractual obligations, and in a manner that permits adherence to this Ethics Code. (Section 9.11, Maintaining Test Security)

CEB 74

Questions for Practitioners

- Do we have an ethical problem?
  - Are we obligated to use the most valid test?
  - Are we obligated to use the most reliable test?
  - Are we responsible for ensuring an appropriate test environment and helping applicants can do their best on an assessment?
- Are the inferences we make about test scores and validity accurate?
  - It depends on what you did and what you say about you did.
    - Did you validate the UIT under the same conditions under which it will be used?
    - Are you disclosing the limitations of the UIT score?
- Are we appropriately protecting an applicant's privacy?

CEB 75

What else can I worry about?

CEB

Other Concerns

- Games
  - How are these scored?
  - How do I standardize the game and take into account the behavior of other players?
  - How can I calculate validity unless the game is scored?
- Branching Tests
  - How are these scored?
- Big Data
  - Is it appropriate to use atheoretical testing procedures?
  - How can I demonstrate that predictive Big Data elements are job relevant?
    - What if they are not?
  - Does Big Data raise ethical and privacy concerns?

CEB

77

Other Concerns

- Brain Imaging
  - Is it ethical? Does it violate privacy concerns?
  - How does it validity compare to other measures?
  - How do applicants react to such measures?
- Dynamic Norms
  - Is the sample size sufficient for the country X function breakout?
  - Is the functional label (e.g., “manager”) sufficiently specific?

CEB

78

Thank You

Nancy T. Tippins
Principal Consultant
864.527.5956
ntippins@cebglobal.com

CEB

CASE STUDIES ON USING TECHNOLOGY AND ASSESSMENT

Dra. Nancy T. Tippins
Vicepresidenta Senior de la compañía
Corporate Executive Board (CEB Inc)

Technology and Assessment - Case Studies

Nancy T. Tippins
CEB

June 24, 2016

CEB
WHAT THE BEST COMPANIES DO

Case Studies

Technology-Enabled Assessments

Nancy Tippins

CEB
WHAT THE BEST COMPANIES DO

- Verbal and Numeric - UIT/ CAT
- General Ability Screen – UIT/CAT
- Video-based Test

CEB

3

Verbal and Numerical

CEB

UIT/CAT

- UIT = Unproctored Internet Test
  - The test is administered over the internet in a location and conditions of the candidate's choosing
- CAT = Computer Adaptive Test
  - Item characteristics are defined using Item Response Theory (IRT) and are aggregated to identify the candidate's ability level

CEB

5

Sample Items

CEB

Verbal Reasoning Management/Graduate Item Bank - Sample 1

Progreso

Instrucciones

Práctica Test de razonamiento verbal: Pregunta 1 de 8

Los directivos y el personal no están de acuerdo en cuanto a los motivos para el absentismo.

Recuerde basar sus respuestas sólo en la información proporcionada en el pasaje.

Existe una marcada discrepancia entre lo que los directivos ven como la causa de las ausencias laborales y las razones citadas por los empleados. Muchas organizaciones en la actualidad experimentan excesivas reducciones de plantilla. Al disminuir los niveles de personal hasta este punto, incluso las cotas más bajas de absentismo pueden poner al resto de los empleados bajo una presión intolerable. Con el tiempo, esto puede aumentar las ausencias relacionadas con el estrés u otros problemas en toda la organización.

- Verdadero
- Falso
- No se puede determinar

Siguiente

EU-SODA-CEW-01:abilityplayer.aspx

Verbal Reasoning Management/Graduate Item Bank - Sample 2

Progreso

Instrucciones

Práctica Test de razonamiento verbal: Pregunta 5 de 8

Hay algunos indicios que indican que los países en vías de desarrollo se benefician de una mayor competencia en el campo de las telecomunicaciones.

Recuerde basar sus respuestas sólo en la información proporcionada en el pasaje.

Los beneficios de la competencia a gran escala en el campo de las telecomunicaciones están bien establecidos en los países desarrollados. No obstante, no resultan ser tan evidentes para los países en vías de desarrollo porque la competencia en sí misma está más limitada. La evidencia disponible apunta al hecho de que la liberalización de los mercados puede ser igualmente beneficiosa para los países en vías de desarrollo, si no lo es más, que para los países que ya están totalmente industrializados, aunque habrá, inevitablemente, algunas víctimas cada vez que se abre un mercado.

- Verdadero
- Falso
- No se puede determinar

Siguiente

EU-SODA-CEW-01:abilityplayer.aspx

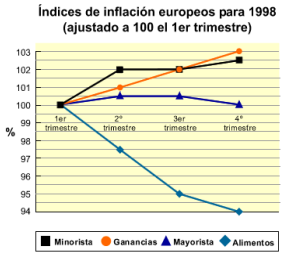
Numerical Reasoning Management/Graduate Item Bank - Sample 1

Progreso

Instrucciones

Práctica Test de razonamiento numérico: Pregunta 1 de 8

Si el índice de inflación en las ganancias del primer trimestre, antes del ajuste (rebasing) era de 120 ¿cuál sería la cifra equivalente en el segundo trimestre?

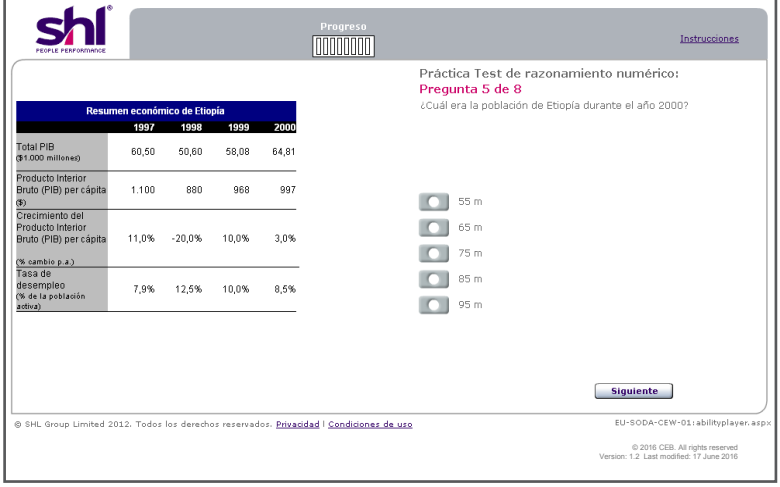


- 121,0
- 121,2
- 121,4
- 121,6
- Ninguna de éstas

Siguiente

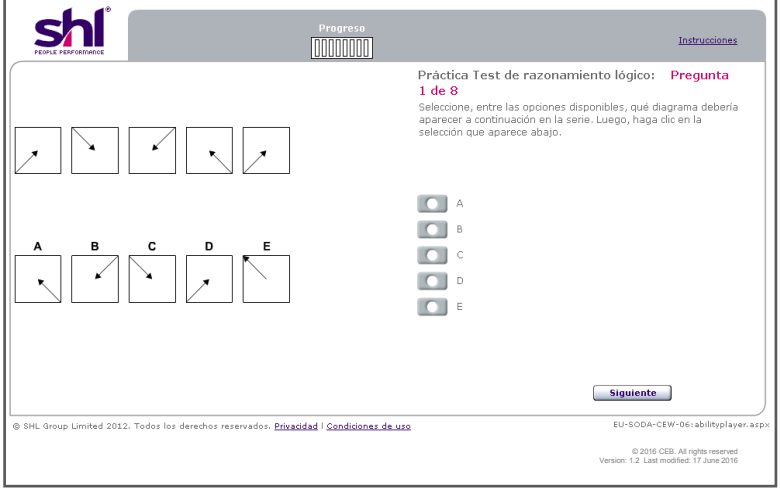
EU-SODA-CEW-01:abilityplayer.aspx

Numerical Reasoning Management/Graduate Item Bank - Sample 2

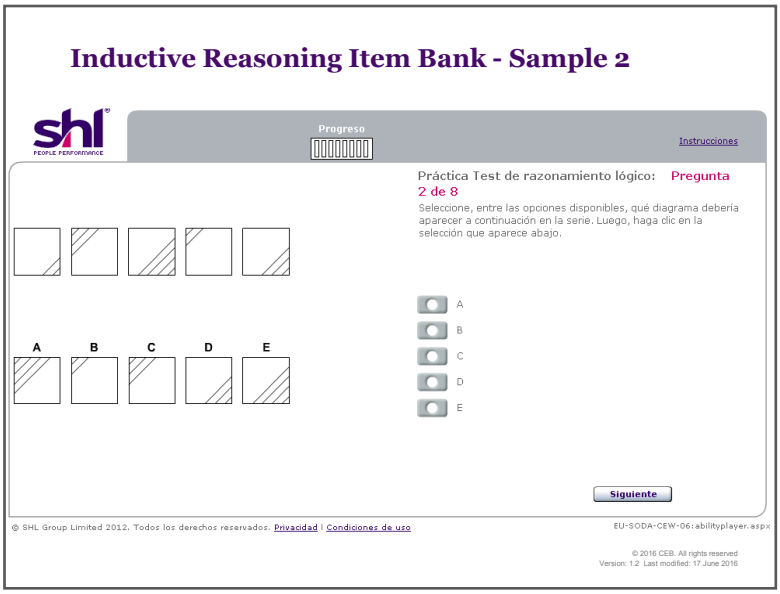


Resumen económico de Etiopía	1997	1998	1999	2000
Total PIB ($1.000 millones)	60,50	50,60	58,08	64,81
Producto Interior Bruto (PIB) per cápita ($)	1.100	880	968	997
Crecimiento del Producto Interior Bruto (PIB) per cápita (% cambio p.a.)	11,0%	-20,0%	10,0%	3,0%
Tasa de desempleo (% de la población activa)	7,9%	12,5%	10,0%	8,5%

Inductive Reasoning Item Bank - Sample 1



Inductive Reasoning Item Bank - Sample 2



Linear-on-the-Fly Test (LOFT) Versions

- Fixed length
- Identical psychometric characteristics across forms

	# Items	Time	# Items	Time
Verbal Reasoning	30	17-19	18	11
Numerical Reasoning	18	17-25	10	14-15
Inductive Reasoning	24	25	7	7

CEB 13

Item Response Theory and Test Bank Construction

- Item characteristics were determined using a 2-parameter model and data from 16,132 participants
  - Θ parameter – the level of ability and difficulty of the items
  - b parameter - point on theta scales where the probability of getting the item correct is 50% (difficulty of the item)
  - a parameter – the slope of the item characteristic curve at the b-parameter for the item (discrimination of the item)
- Items screened on the following:
  - Sensitivity, i.e., give offense to minority groups
  - Low a-parameters (low discrimination)
  - Extreme b-parameters (less than -3 and greater than +3)
  - Long response times (e.g., 2+ minutes)
  - Distractors with positive item-total scores

CEB 14

Scoring of Computer Adaptive Test (CAT) - Verify

- Estimate θ on initial set of items for which a and b parameters are known
- Using initial θ estimate and item properties, calculate the expected probability of getting the item correct
- Calculate the difference between the candidate answering correctly and the expected probability of the candidate getting the item correct
- Sum the differences; standardize the differences, add standardized difference to initial θ estimate
- Repeat the cycle until the difference between the value of the initial θ and the new θ are trivial

CEB 15

Reliability

	IRT	CTT	Verbal	Numeric
Score Accuracy	SE (θ)	SEM		
	(depends on the properties of the questions included and the range of θ)	(Constant across all scores)		
Variation across Forms (100 forms)	Alternate Forms Coefficient	Alternate Forms Coefficient	0.81 (median – mgmt. & grad) 0.78 (median – Sup & Op)	0.83 (median – mgmt. & grad) 0.94 (median – Sup & Op)
Test-Retest (test/verification test)	Test-Retest Coefficient	Test-Retest Coefficient	0.72	0.70

CEB 16

Meta-Analytic Criterion-Related Validity

	Verbal	Numerical
Number of Studies	5	7
Total Sample Size	548	760
Average Sample Size	110	109
Range of Observed Validities	0.21 to 0.43	0.11 to 0.34
Variance in Observed Validity (a)	0.01	0.00
Sampling Error across Studies (b)	0.01	0.01
True Variance in Validities (a-b)	0.00	-0.01
Weighted Mean Operational Validity	0.50	0.39

CEB 17

Construct Validity

Measure	VIR	GMA	Ravens	Numerical	Verbal
VIR	0.770	0.538	0.558	0.315	0.391
GMA	0.655	0.875	*	0.374	0.390
Ravens	0.690	*	0.850	0.395	0.454
Numerical	0.404	0.450	0.483	0.788	0.248
Verbal	0.496	0.464	0.548	0.311	0.808

- VIR = Verify Inductive Reasoning
- GMA = Graduate and Management Assessment - Abstract
- Ravens = Ravens Progressive Matrices
- Diagonal = reliabilities
- Above the diagonal = uncorrected
- N=49-109

CEB 18

Questions When Verification Test Score Is Flagged

- What was the candidate's physical condition? How did the condition differ from that when the original test was taken?
- Were there reasons the candidate was unable to focus while taking the verification test?
  - Were there distractions in the Verification Test?
  - Were there physical or psychological reasons why the candidate could not focus?
- Did the candidate attempt all of the questions in the Verification Test?
- Did the candidate attempt to the practice items?
- Why does the candidate think the two scores differ?

CEB 19

Discussion Questions

- What is the appropriate criteria for evaluating the difference between the test and the verification test?
- At what point is a test score disqualified?
  - Now what?
- What are the advantages of LOFT compared to CAT?

CEB 20

General Ability Screen

CEB

General Ability Screen

- Purpose screen out unqualified entry-level candidates
- Assesses general cognitive ability through the following:
  - Numerical Ability
    - Calculations
    - Charts and graphs
  - Inductive Reasoning
    - Prediction
    - Missing step
  - Verbal Ability
    - Reading comprehension
  - Deductive Reasoning
    - Syllogisms
    - Games
    - Logical arguments
- 12 questions; 10 minutes
  - Globally adaptable
  - Narrow range of difficulty: -2.0<b<0
  - Discrimination (a) > 1.0
  - Guessing : <0.25 for deductive, inductive, and quantitative; <0.30 for verbal
  - 80% of population can complete 12 items in 10 minutes
  - Reliability (i.e., square root of the correlation between true score and estimated score) = .87

CEB 22

Criterion

- Performance Ratings
  - Using Numbers
  - Writing and Reporting
  - Making Rational Judgments
  - Analyzing
  - Learning and Researching
  - Creating and Innovating
  - Planning and Organizing
  - Learning
  - Job Knowledge
  - Productivity
- Key Performance Indicators:
  - Mathematical Operations
  - Oral and Written communications
  - Problem Solving
  - Generating Solutions and Ideas
  - Gathering Information
  - Customer Problem Solving
- Global Ratings
  - Re-hireability
  - Match between abilities and job requirements
  - Employee productivity level

CEB 23

Meta-Analytic Validity

Criterion	Number of Studies (k)	Sample Size (N)	Observed Correlation (r)	Estimated Operational Validity (ρ)[1]	Percent of Variance Accounted for by Sampling Error	Credibility Interval Lower Bound	Credibility Interval Upper Bound
Performance Area Composite	9	1,403	.13	.17	100	.17	.17
Cognitive Composite	7	1,000	.19	.24	100	.24	.24
Global Performance Composite	10	1,566	.05	.07	100	.07	.07
Total Composite	8	1,236	.10	.13	100	.13	.13

[1]Correlations in the Estimated Operational Validity column have been corrected for criterion unreliability.

CEB 24

Subgroup Differences

Group	Gender		Age		Racial/Ethnic Group						
	Male	Female	<40 years	≥40 years	White	Black/ African American	Hispanic	Asian	Two or More Races	American Indian	Native Hawaiian/ Other Pacific Islander
N	11,989	27,308	11,498	5,141	26,628	5,903	2,349	1,828	1466	243	241

Female*	≥40**	Black/ African American***	Hispanic***	Asian***	Two or More Races	American Indian***	Native Hawaiian/Other Pacific Islander
-0.17	-0.54	-0.66	-0.43	-0.14	-0.19	-0.41	-0.45

*Referent group is Male.
**Referent group is <40 years old.
***Referent group is Caucasian.

Effect Sizes
Small to Medium: >|0.2| to ≤|0.5|
Medium to Large: >| 0.5| to ≤ |0.8|
Large: ≥ |0.8|

Cohen (1988).

CEB

25

Discussion Questions

- How much speed is needed to reduce cheating significantly?
- What is the incremental validity of more in-depth tests over the short screening test?

CEB

26

Agenda

UIT/CAT
1) Verbal
2) Numeric
3) General Ability

Technology-Enabled Assessment

Technology-Enabled Assessment

Electronic SJIs

CEB

27

Video-Based Testing (VBT) at U.S. Customs and Border Protection

- Job Duties – duties associated with passport control, customs, immigration, agricultural, and anti-terrorism.
- KSAOs (knowledge, skills, abilities, and other characteristics)
  - Logical reasoning
  - Math skills
  - Interpersonal skills
  - Emotional maturity
  - Cooperativeness/sensitivity to the needs of others
  - Judgment/decision making
- VBT measures interpersonal skills, emotional maturity, cooperativeness/sensitivity, judgment/decision making
  - Previously measured via structured interview

CEB 28

Current VBT

- Video-based scenarios
  - Requires applicants to respond to situation rather than say what they would do
  - Provides a realistic job preview
  - Reduced staff time to conduct interviews
    - Two supervisors for 15 minutes (review of responses only) vs. three supervisors for one hour (interview)
  - Reduced interviewer travel expenses – VBT can be deployed anywhere with CBP personnel
  - Eliminates extensive note taking required in interviews because VBT is recorded

CEB 29

Eight Development Steps

1) Conduct job analysis to identify critical tasks and KSAOs
2) Collect critical incidents that arise on the job
3) Identify critical incidents that are amenable to testing
4) Create brief scenarios from critical incidents
5) Conduct pre-production activities
   1) Draft scripts
   2) Identify and audition actors
   3) Procure props
   4) Create filming schedule
6) Produce video
7) Ask SMEs to rate filmed scenarios on importance, difficulty, frequency, and competency coverage and define probable applicant response to create benchmarks
8) Piece scenarios into multiple versions of VBT matched b difficulty and competency coverage

CEB 30

Administration of VBT

- VBT is administered by trained test administrator
  - Provide oral and written instructions
  - Starts VBT tape and leaves
    - Introduction
    - 2 sample scenarios
    - 8 evaluated scenarios
- Applicant has 45 seconds to respond to each scenario
  - Response is recorded for later evaluation
- VBT is scored by panel of raters
  - Trained for 1.5 days
- ~15K VBTs are administered per year

CEB 31

Scoring the VBT

- Two raters
- Phase I
  - Independently score each response on the competency measured using a 3-point scale
  - Sum ratings across 8 scenarios for each competency
  - Compare to cutoff score for each competency
- Phase II
  - Share individual scores for each scenario
  - Reach consensus
- Applicant must meet the cutoff score for each competency
- Agreement
  - Pass/Fail status = agree 96-98%
  - Overall Pass/Fail status – agree 95%

CEB 32

Validity

- Content Validity
  - VBT is designed to measure key competencies identified in traditional job analysis
  - Critical incident technique used to identify job relevant scenarios
  - Scenarios reviewed throughout development process to ensure relevance to job
  - SMEs rated importance, frequency and competency coverage
- Construct Validity
  - Based on relationship with other measures
    - VBT and
      - Cognitive tests (*t=3.024, p = .003 , d=.29)*
      - Logical reasoning (*t=2.953, p = .003, d=.35)*
      - Writing (*t=2.189, p = .029 , d=.26)*
      - Low fidelity simulation (*t=2.454, p = .016* , d=*.42)*

CEB 33

Costs

- $59 administration and scoring
  - Half the $137 for the structured interview
- Costs
- Development
  - SME Panel Travel
  - Test Developer Time/Salary
  - Extra Reviews of Scenarios
  - Video Production
    - Script writing
    - Salary and expenses for video production staff
    - Travel
    - Pay for actors
    - Post-production editing costs
    - Props and wardrobe
- Implementation
  - Test Production and Distribution
  - Administration Equipment
  - Supplies for Administering and Rating the VBT
  - Rater and Administrator Time
  - Program Management Time

The average utility for a video-based assessment tool (using only significant validities) was $109,183,823.

CEB 34

Applicant Reactions

1. How would you rate your performance on this test?
   a. I performed exceptionally well on this test. 11%
   b. I performed above average on this test. 39%
   c. My performance on this test as about average. 48%
   d. I performed below average on this test. 2%
   e. I performed poorly on this test. 0%
2. How comfortable were you in responding to the scenes on the TV monitor?
   a. I was very comfortable responding to the scenes on the TV monitor throughout the entire test. 21%
   b. I became comfortable responding to the scenes on the TV monitor after the practices scenes were given. 32%
   c. I became comfortable responding to the scenes on the TV monitor after responding to a couple of the real test scenes. 44%
   d. I was not at all comfortable responding to the scenes on the TV monitor throughout the test. 3%

CEB 35

Applicant Reactions

3. How sufficient were the instructions for this exam?
   a. The instructions were sufficient for responding to the scenes on this exam. 95%
   b. The instructions were somewhat sufficient, but I could have used more instruction prior to responding to the scenes on this exam. 4%
   c. The instructions were not sufficient, resulting in confusion. 0%
4. How would you rate the forty-five-second response time at the end of each scenario?
   a. Too long. 24%
   b. Just right. 75%
   c. Too short. 1%

CEB 36

Applicant Reactions

5. What did you like most about the video-based test?
   a. Depicted real-life situations. 19%
   b. Gave a realistic job preview. 15%
   c. Miscellaneous (categories that were only mentioned by two or fewer respondents). 11%
   d. Was alone in the room and didn't have to appear before a panel. 7%
   e. Instructions were good/detailed. 7%
   f. Efficient/fast/concise. 3%
   g. Test was good/fair. 2%

CEB 37

Applicant Reactions

6. What did you like least about the video-based test?
   a. Provided a positive comment/didn't dislike anything about the VBT. 22%
   b. Miscellaneous (categories that were only mentioned by two respondents or fewer). 18%
   c. No feedback from characters/actors on TV monitor. 15%
   d. VBT Format was hard to get used to. 9%
   e. Response time was too long. 8%
   f. Nervous on camera. 4%
   g. Impersonal (no interaction with a real live person). 3%

CEB 38

Discussion Questions

- How does the validity of the VBT compare to a more abstract test such as a measure of cognitive ability?
  - What is the cost/benefit ratio of each?
- How does one maintain a test like the VBT?
- Is a cheating a significant threat here?
- What are the "comparative" applicant reactions?

CEB 39

Thank You

Nancy T. Tippins
Principal Consultant
864.527.5956
ntippins@cebglobal.com

CEB

EL RECLUTAMIENTO DE PERSONAL MEDIANTE RECURSOS WEB

Antonio L. García-Izquierdo
Profesor Titular de la Facultad de Psicología de la Universidad de Oviedo y Director de la Cátedra Universitaria Asturias Prevención

El reclutamiento de personal mediante recursos web

Cátedra de Modelos y Aplicaciones Psicométricos:
"Nuevas tecnologías en evaluación y medición"
24 de junio 2016

CÁTEDRA ASTURIAS PREVENCIÓN

Universidad de Oviedo

Antonio León García Izquierdo
Profesor Titular Universidad de Oviedo
Área Psicología Social
IP Proyecto https://espuma.uniovi.es/
Director Cátedra Asturias Prevención
angarcia@uniovi.es

Índice de contenidos

2

- **Reclutamiento y Selección de personal: Ajuste Individuo-Organización**
- **Las redes sociales**
  - **Características y uso**
- **El Reclutamiento y el e-Reclutamiento (e-R)**
  - **Ventajas y desventajas**
- **Los Biodata en el e-R**
- **Estudio empírico sobre (e-R): análisis de contenido de los formularios web. Controversia sobre Biodata, Justicia Organizacional y aspectos legales**
- **Conclusiones**
- **Investigación futura**

Actualidad de los Social Media

3

- El estudio de la *Society for Human Resource Management* muestra que el 56% de las empresas usan los SM para encontrar candidatos, cuando fue de un 34% en 2008. Además, en 2011 más del 25% of comprobaron los perfiles de los candidatos en los SM.
- El portal de empleo Jobvite (2013), señala que el **93% de** los **reclutadores** son propensos a conocer el perfil de los candidatos en los social media, y que **el 43% ha reconsiderado** su decisión basándose en ello.
  - La información se evalúa según se perciba como positiva o negativa. Según el estudio de (Careerbuilder.com, 2009) el 35% de los reclutadores dijeron rechazar a un canidato por la información negativa, mientras que sólo el only 18% de contrató por la información positiva.
- En un estudio realizado con 30,000 graduados, estudiantes y profesionales junior en toda Europa, *casi el 100 por cien estaban dispuestos a interacionar con las empresas online, preferentemente mediante LinkedIn 48 %, seguido de Facebook 25 %* (Potentialpark, 2011).
- Nigel Wright Recruitment (2011), encontró que más de la mitad de los buscadores de empleo del Reino Unido busca infformación en las redes sociales

Desarrollo de las perspectivas sobre RyS

4

- La **ortodoxia** en la selección de personal (Guion, 1976) se ha denominado el **paradigma dominante** y está basado principalmente en la Psicometría (Herriot, 1992).
- No osbtante, desde hace 50 años la investigación académica comenzó a reconocer el Reclutamiento y la Selección (RyS) como un proceso de ***interacción social motivacional*** (Derous, van der Velde, & Born, 2011).
  - Ya en 1965, Schein adelantó el impacto que la ***tecnología*** tendría en este campo.
- Los **cambios sociales** han llevado a los especialista en (RYS) a afrontar como la evaluación y predicción del desempeño en un entorno dinámico y de cambio constante, por lo que las perspectivas han cambiado desde el paradigma tradicional clásico al paradigma social, como una ***negociación bilaterial*** como ya desarrolló Herriot (1989, 1992) en el que la ***perspectiva del candidato*** **es mucho más tenida en cuenta**

5

- **Ajuste y** C***omunicación/ negociación bilateral:***
- Propicia nuevos temas de investigación que tiene en cuenta una mayor **diversidad**, **justicia**, **imparcialidad**, selección para el cambio, etc.

  Consecuentemente, la perspectiva tradicional y las social se pueden considerar complementarias antes que enfrentadas (Derous & De Witte, 2001).
- Siguiendo el modelo **Attraction-Selection-Attrition, (ASA)** (Schneider, 1987) **y a** Wanous (1992): es un ***ajuste bilateral*** Individuo-Organización, en el que se da un proceso de **toma de decisiones**:
  - Intereses, necesidades, personalidad del individuo y la capacidad de la organización para satisfacerlas y viceversa.
    - **Consecuencias** del **buen Ajuste** en la investigación: mayor satisfacción y mayor compromiso
    - **Consecuencias del Desajuste** lleva a baja satisfacción y estrés (Kristof-Brown & Guay, 2011), absentismo y rotación interna o externa (salida de la organización).
- Según Porter, Lawler y Hackman (1975): elección *f (atractivo x factor de realidad*), que le permite al sujeto ponderar las posibilidades que tiene de conseguir dicho puesto.

  En muchos casos, no intentan conseguir los trabajos que les resultan más atractivos **porque juzgan que sus posibilidades** de conseguirlos son nulas o muy reducidas.

Relación Individuo-Organización y la web

6

Individuo: *atractivo para las organizaciones adaptando su comportamiento* a lo que piensa que son los deseos de la organización, lo que da lugar a

"Impression management", **pero también selecciona** organizaciones a partir de la información que el sujeto acumula y que obtiene de muy diversas fuentes: anuncios de reclutamiento, publicidad general de la empresa, a través de empleados, etc., y cada vez más de lo que aparece en la **web** (información de la web oficial y también de comentarios de usuarios, blogs, etc.

Esto fomenta ***la autopromoción y la autopresentación narcisista*** (Mehdizadeh, 2010), además de la necesidad **de *popularidad*** (Christofides, Muise, & Desmarais, 2009) Según (Magnuson & Dundes, 2008), hombres y mujeres adoptan autopresentaciones distintas, roles clásicos de masculinidad y femineidad siendo los ***varones más instrumentales*** *y menos emocionales, y las* ***mujeres*** *se las ha considerado* ***más expresivas*** (con más fotografías, por ejemplo).

Estudio por LinkedIn (Linkedin User Statistics and Demographics, LUSD) afirma que una persona que **añade una foto profesional a su perfil de usuario tiene 14 veces más probabilidades** de ser contratado que una persona que no usa foto. La red social también ha revelado que los **perfiles que tienen una lista de habilidades y competencias** (skills) son vistos 13 veces más que aquellos perfiles que no las han incluido (LUSD, 13/8/2015).

Organización: También trata **de atraer** individuos con el propósito de obtener un número suficiente de candidaturas idóneas; y al mismo tiempo **selecciona** a los individuos con el propósito de identificar a aquéllos que mejor se adecúen a los requisitos definidos por ésta.

Debido a **la escasa información que** normalmente disponen los candidatos sobre los puestos de trabajo ofrecidos, la **decisión** inicial de solicitar una vacante está fuertemente basada en **impresiones generales sobre el atractivo de la organización**.

Redes sociales y Reclutamiento 2.0

7

- El término **Web 2.0 o Web Social**[1] comprende aquellos sitios web que facilitan el compartir información, la interoperabilidad, el diseño centrado en el usuario y la colaboración en la World Wide Web.
  - Crean un *ciberespacio* propio sin delimitación espacial, siendo las restricciones las culturales y las idiomáticas.
  - Originan un *cibertiempo* propio, ya que permite participar en un tiempo sincrónico e ininterrumpido, aunque abre también tiempos diferidos en los que los usuarios intervienen cuando desean sin sujeción a la simultaneidad.
  - Es una conversación abierta, en la cual se dialoga, se comparte, se interactúa, se intercambia información. El usuario genera datos, todo aquello que comparte, clasifica y en aquello en lo que participa. Da lugar a *otras relaciones* diferentes a las de la vida cotidiana.
  - Publicidad viral, la publicidad la desarrolla el propio usuario mediante sus redes sociales e incluso blogs, siendo por tanto mucho más efectiva, menos costosa y con poder de fidelización sobre marcas.
  - **Permite** (Boyd y Ellison, 2008):
    - Crear un **perfil público o semi-público** en un sistema cerrado
    - Articular **un listado restringido de usuarios** con los que comparte información, y ver la información que emiten quienes comparten el sistema
    - Participación activa **creando una personalidad "*online*"**

Ventajas uso SM:

8

- Los reclutadores pueden captar los aspectos de personalidad de los candidatos en sus perfiles. Según Caers & Castelyns (2011), los reclutadores participantes afirmaron que eran capaces de detectar el grado de responsabilidad, de estabilidad emocional y de madurez en los perfiles.
  - El número de contactos de un candidato, la ortografía, las recomendaciones o la cantidad de información personal que se revela son algunos de los elementos que se tienen en cuenta a la hora de evaluar un perfil en LinkedIn (Zide, Elman y Shahani-Denning, 2014).
  - Bajo coste (Jacobs, 2009), parecen fiables para los usuarios (Kluemper and Rosen, 2009).

  Según (Kroeze, 2015: Smith and Rupp, 2004)
  - Reclutadores lo usan para comprobar la veracidad de los CV y conocer su capital social (Reiners & Alexander, 2013)
  - Aumentar la información que no aparece en el CV
  - Pueden acceder con mayor facilidad y rapidez a candidatos internacionales.
  - También pueden filtrar y cruzar información.
  - Los formularios web se pueden estandarizar y simplificar el proceso, e incluso automatizarlo (Compton et al., 2009) (Torrington et al., 2004).
  - La carga de trabajo se ha trasladado al candidato, por lo que éstos deben asegurarse de mantener sus plataformas de medios sociales actualizadas y con contenido atractivo y preciso.
  - Realizar entrevistas a distancia
  - Personalizar la comunicación con los candidatos
  - Mejora de la comunicación y actualización de contenidos de los procesos
  - Posibilidad de soporte en varios idiomas

Desventajas uso SM

9

- Los reclutadores pueden:
  - **Fijarse** en información **falsa**, o que se puede malinterpertar por el contexto y puede llevar a un rechazo impulsivo (Brown & Vaughn 2011, p. 220-221). Un estudio de *ADP Screening and Selection Services* encontraron que los candidatos falsifican la información en los formularios, como los las evaluaciones de su desemepño, formación, recomendaciones, etc. (Levashina, 2009).
  - **D**ejarse influenciar por información que no está relacionada con el rendimiento futuro del candidato, como por ejemplo tatuajes (Seiter & Hatch, 2005) y otra información personal revelada en el perfil (amistades, preferencias, actividades de ocio, etc.). Esto mismo puede tener consecuencias legales, como ser fuente de discriminación o el uso de usar información irrelevante (Segal, 2014).
  - Segal, (2014) También es probable que la información que se sube a la red sea imprecisa (Suder, 2014)
- En el estudio de Madera (2012) se percibe a las organizacioens que usan estos métodos como menos justas e imparciales, lo que puede llevar a la pérdida de candidatos válidos.
- Asimismo, la información que se vierte en la red es muy variada lo que no permite comparar a diferentes candidatos de manera fiable.

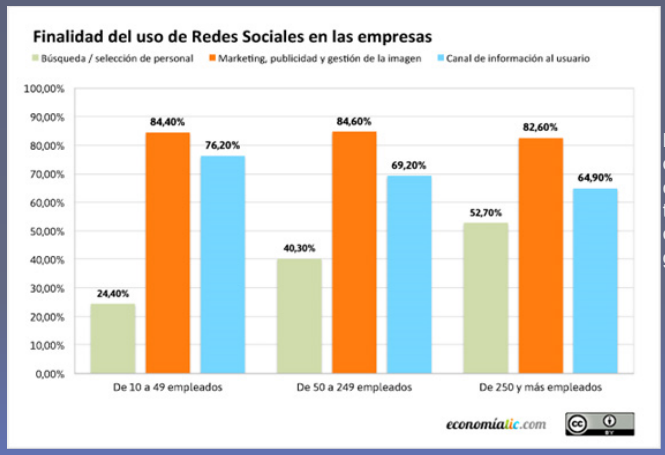
Controversias uso de los social media en el reclutamiento

10

- Según Drouin, O'Connor, Schmidt y Miller (2015) en una muestra de 448 estudiantes, en torno al 60% estaba **en *contra del uso de los social media*** *para decisiones sobre empleo.*
  - Facebook permite **una configuración de privacidad**, pero las personas no siempre se protegen eficazmente (Chamarro, Bertrán, Oberst & Torres, 2015).
  - *Paradójicamente,* los millenial exigen privacidad a los empleadores que saben pueden ver sus perfiles pero eso no les restringe la cantidad y tipo de información que suben a la red (Sánchez, Levin y del Riego, 2012)

Uso de redes sociales en España (2012) -

11



Búsqueda de candidatos en color gris

Reclutamiento

12

- Es el primer paso en la Selección de Personal (SP), y se puede definir como el proceso para ***atraer*** individuos competentes para ocupar un puesto de trabajo (Schmitt & Chan, 1998).
- Más específicamente, Stone, Lukaszewski, Stone-Romero ay Johnson (2013) señalan que su propósito es "proporcionar a la organización de un conjunto de candidatos ***que se ajusten*** a las vacantes en términos de sus conocimientos, habilidades, capacidades y otros atributos" (p. 51).

13

e-Reclutamiento (e-R)

- Es una forma de reclutamiento externo y en la actualidad, Internet es la principal fuente de reclutamiento en todo el mundo (Pfielffelmann, Wagner & Libkuman, 2010).
- Según *el Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información* (ONTSI), en enero de 2013, el **71. 6% de las empresas con más de 10 empleados** tienen su propia web, y el **21.9%** la usa para publicar sus ofertas de trabajo o recibir propuestas de candidatos.
- Además de propocionar información sobre la empresa contratante, un elemento importante es la disponibilidad de formularios online para optar a los puestos de trabajo (Sylva & Mol, 2009)

14

Ventajas y limitaciones del e-R

- *Ventajas*:
  - Según Cober, Brown, Keeping and Levy (2004), los que buscan empleo son activos y la interacción es más amigable (sonido, imágenes, etc.).
  - El uso de la web propia permite dar más información y más pesonalizada (Lievens & Harris, 2003), lo que permite atraer a futuros colaboradores (Cober, Brown, Blumenthal & Levy, 2001) una mejor imagen (Thomson, Braddy & Wuensch, 2008), y ahorro de costes (Sylva & Mol, 2009).
  - Conllevan menores costes, un menor tiempo para la recogida de datos, y acceso las 24 horas del día, independienemtente de donde se encuentren. (Sylva & Mol, 2009; Viswesvaran, 2003).
  - Añaden información que no se ha generado para el reclutamiento, tanto a nivel personal como profesional, así que se puede recabar información complementaria que los candidatos aportan sin ser conscientes de su propósito. (Broughton, Foley, Ledermaier y Cox, 2013)
- *Limitaciones*: Restricción de candidatos (Boehle, 2000), y algunos pueden ser exlcuidos por aspectos demográficos y por la barrera digital (McManus & Ferguson, 2003).

15

Los *Biodata en el reclutamiento*

- Son fáciles de recoger y poseen una aceptable validez predictiva (between .30 and .40) con diferentes criterios (Allworth & Hesketh, 2000).
- Los candidatos no suelen falsearlos (Schmitt & Kunce, 2002).
- Limitaciones: en ocasiones pueden abordar aspectos no legales (Furnham, 2008) y parece que pueden provocar una mayor impacto adverso del que se creía (Bobko & Roth, 2013).
  - Por ejemplo, incluir contenido donde se sugiera la preferenciar por determinado género cunado no se relaciona con el rendimientopuede ser un promer paso hacia la discriminación (Stone-Romero & Stone, 2005). Asimismo, la inclusión de contenido invasivo o inapropiado puede provocar la autoexclusión de candidatos potencialmente válidos (Truxillo et al., 2004) y generar reacciones negativas (Bauer et al., 2006).

Estudio e-R en España (2005-2009), publicados en 2010 y 2015 *(IJSA, 2010-* Science-practice Gap in E-Recruitment Antonio L. García-Izquierdo · Herman Aguinis · Pedro J. Ramos-Villagrasa; *y JWOP-2015,* E-Recruitment, gender discrimination, and organizational results of listed companies on the Spanish Stock Exchange, Antonio L. García-Izquierdo · Pedro J. Ramos-Villagrasa · Ana M. Castaño)

16

- ***Internet Recruitment Process Model*** (Cober, Brown, Blumenthal, & Levy, 2001), based on websites information
- *Indirect discrimination source*: perspective of applicants, the information collected through online job applications could be used to make decisions that are procedurally unfair regardless of the actual outcome of the hiring process (Bell, Ryan, & Wiechmann, 2004). For example, information **such as place of birth, nationality, marital status, age, and disability status has high potential to be used for procedurally unfair decisions and also motivate applicants to initiate litigation** (Goldman, 2001). **intrusiv**e (Anders on, 2003) and constitute an **invasion of privacy** (Eddy, Stone, & Stone-Romero, 1999).
- Empirical evidence also suggests that collecting information seen as intrusive and violating an applicant's privacy leads to negative applicant reactions (Bauer et al., 2006; Cascio & Aguinis, 2011, Stone-Romero & Stone, 2005). In short, there is abundant scholarly research providing evidence that **e-recruitment practices that are potentially illegal and perceived to be unfair, ¡intrusive, and invasive of applicants' privacy** are associated with **negative reactions ranging from negative perceptions and emotions to actually initiating legal action against the recruiting firm.**

Fairness and reactions to personnel selection methods

17

- Anderson, Born, & Cunningham-Snell (2001, p. 205) state that applicant's reaction to test taking are more favourable when these factors are met:
  - "(i) more **job-relevant**,
  - (ii) **less personally intrusive**,
  - (iii) not **contravening candidate procedural or distributive justice expectations**, and
  - (iv) **allowing the candidate an opportunity to meet** in person with selectors."

Objective

18

- To synthesize application forms and discourse contents in order *to extract discriminatory content against women*, in listed companies' Web sites.
  - Specially:
    - Direct discrimination- Differential treatment
    - Indirect discrimination- Adverse impact

Theoretical and legal grounds in the study

19

- Stereotypes and Decision-making in Management
  - Web sites:
    - Highhouse & Hoffman (2001): Decision-making and job-seeking perspective, cues, signal and heuristics: to be analyzed in Web sites discourses
  - Application forms:
    - Stone-Romero & Stone (2005): Decision-making choices and sterotypes perspective: to be analyzed in companies´ application forms
    - Gilliland (1993): opportunity to perform rule of procedural justice
    - Stone & Stone (1990): Organizational Privacy Theory (information control and invasion perception)
- Legislation in force, specially 3/2007 Spanish Organic Act on Equality between Men and Women

EEO legislation in Europe

20

- Gender equality is embedded in the Treaty establishing the European Union
- EU Member States approved a European pact for gender equality at the European Council meeting on March 2006
  - Communication: A Roadmap for equality between women and men 2006-2010 SEC(2006)
    - Two of the priority areas are: *equal representation in decision –making and promotion of gender equality*, *eliminating gender stereotypes in society* (labor-market), and *better governance* at all levels, EU institutions, Member States, parliaments, social partners and civil society

Method

21

- Web sites' content analysis (need proxies to assess mental models as executives are difficult to access to, and if so they probably neglect they use such models)
- Comparison of discriminatory gender related items before (2005) and after (2009), the 3/2007 Constitutional Act

22

Sample and Procedure

- Initial 186 organizations sites quoted on Spanish Stock Exchange (Madrid)
  - Final sample n=110 (due lack of information or pertaining to consortia)
- Data collection 2005 and 2009
  - Software: TexStat 2.0; QSRNVIVO 2.0
- Content Analysis: Iterative Codification
  - Unit analysis: Web sites' HR sections during march and may in 2005 and 2009
    - Discourses
    - Application forms
  - Code structure and categories:
    - Multiple coders (kappa index=.92)
  - Data analysis

23

Results: Discourses

- Discourses (n=75)
  - Categories: **Personnel** (e.g. People), **Personal requirements** (e.g. Effort), **Processes in the company** (e.g. Training), **Processes related to the company** (e.g. Strategy)
    - More Freqeuents: Work (69.33%), Professional (66.67%), Team and People (65.33%)
  - Discourse's organization:
    - Number of employees (36.00%), Industry (30.67%), Mission/Objective (14.67%), Company's history 12%
    - To whom is directed (professions and required profiles). Worker's and masculine (97.33%), ***only 2.67% both genders***.
    - Topics about EEO or WorkLifeBalance (4.00%)
    - What they offer (centred on values), training (62.70%), globalization (22.70%), teamwork (20.00%), mobility (10.60%),
    - What they are looking for (centred on **values** and **young candidates** (22.67%)
    - **Messages for attracting candidates**

24

Results: Application forms 1

- N=76.
- 108 Categories in 3 sets:
  - **Personal data** >40%: Age / date of birth, Identity card, Nationality, Sex / gender*, Marital status *, Place of birth, Passport
  - **Availability** >40%:, Mobility and Willingness to travel
  - **Other data** >40%: Driving license, hobbies, etc.
- Gender bias categories (2005- 2009)
  - Sex/gender (57.89%- 66.67%)
  - Marital status (51.31%- 45.45%
  - Number of children (10.53%- 12.12%)
  - Military service (3.95%- 3.03%)

Table 1. Companies classified by industry and Web site sections: frequency and percent

25

Sections	Industry						
	Raw materials, manufacturing and building	Consumer goods	Financial services and real estate	Consumer services	Oil and energy	Technology and telecommunications	Total
Application form	4 (9.76%)	9 (25.00%)	1 (2.86%)	3 (15.00%)	3 (18.75%)	1 (12.50%)	21 (13.46%)
Discourse	8 (19.51%)	2 (5.55%)	1 (2.86%)	1 (5.00%)	5 (31.25%)	1 (12.50%)	18 (11.54%)
E-mail	4 (9.76%)	1 (2.78)	1 (2.86%)	2 (10.00%)	0 (0.00%)	0 (0.00%)	8 (5.13%)
Application form and discourse	12 (29.27%)	9 (25.00%)	10 (28.57%)	8 (40.00%)	6 (37.50%)	2 (25.00%)	47 (30.13%)
Application form and e-mail	2 (4.87%)	2 (5.55%)	1 (2.86%)	0 (0.00%)	0 (0.00%)	1 (12.50%)	6 (3.85%)
Discourse and e-mail	2 (4.87%)	1 (2.78%)	1 (2.86%)	3 (15.00%)	0 (0.00%)	1 (12.50%)	8 (5.13%)
Application form, discourse and e-mail	0 (0.00%)	1 (2.78%)	0 (0.00%)	1 (5.00%)	0 (0.00%)	0 (0.00%)	2 (1.28%)
Does not have any	9 (21.95%)	11 (30.55%)	20 (57.14%)	2 (10.00%)	2 (12.50%)	2 (25.00%)	46 (29.48%)
Total	41 (100.0%)	36 (100.0%)	35 (100.0%)	20 (100.0%)	16 (100.0%)	8 (100.0%)	156 (100.00%)

Note. $N = 156$.

Table 2. Categories in discourses

26

Categories used by ≥40.00%		Categories used by ≤10.00%	
Work	52 (69.33%)	Efficiency	7 (9.33%)
Development	50 (66.67%)	Flexibility	7 (9.33%)
Professional	50 (66.67%)	Learning	7 (9.33%)
People	49 (65.33%)	Quality of life	7 (9.33%)
Team	49 (65.33%)	Creativity	6 (8.00%)
Company	41 (54.60%)	Integrity	6 (8.00%)
Employees	39 (52.00%)	Join	6 (8.00%)
Group	39 (52.00%)	Mission	6 (8.00%)
Resources	39 (52.00%)	Attract	5 (6.67%)
Training	38 (50.67%)	Driving force	5 (6.67%)
Human	38 (50.60%)	Employee social benefits	5 (6.67%)
Personnel	36 (48.00%)	Mobility	5 (6.67%)
Professional development	34 (45.33%)	Participants	5 (6.67%)
Values	30 (40.00%)	Philosophy	5 (6.67%)
		Promote	5 (6.67%)
		Environment	4 (5.33%)
		Ethics	4 (5.33%)
		Honesty	4 (5.33%)
		Risk assessment and prevention	4 (5.33%)
		Transparency	4 (5.33%)
		Disability	3 (4.00%)
		Discrimination	3 (4.00%)
		Equal opportunities	3 (4.00%)
		Professional profile	3 (4.00%)
		Work and life balance	3 (4.00%)
		Loyalty	2 (2.67%)
		Recruitment	2 (2.67%)
		Discretion	1 (1.33%)
		Geographical mobility	1 (1.33%)
		Open mind	1 (1.33%)

Note. $N = 75$.

Table 3. Categories in applications forms

27

Set	Categories in application forms	Year 2005 (n = 76)	Year 2009 (n = 66)
Personal data	Age / date of birth	67 (88.16%)	57 (86.36%)
	Identity card	60 (78.95%)	60 (90.91%)
	Nationality	45 (59.21%)	41 (62.12%)
	Sex / gender*	44 (57.89%)	44 (66.67%)
	Marital status *	39 (51.31%)	30 (45.45%)
	Place of birth	38 (50.00%)	42 (63.64%)
	Passport	33 (43.42%)	33 (50.00%)
	Photograph	15 (19.74%)	17 (25.76%)
	Green card	15 (19.74%)	15 (22.73%)
	Number of children *	8 (10.53%)	8 (12.12%)
	Work license	6 (7.89%)	6 (9.09%)
	Disability	5 (6.58%)	6 (9.09%)
	Relatives working at organization	5 (6.58%)	4 (6.06%)
	Military service* (sin efecto obligatorio desde 2001)	3 (3.95%)	2 (3.03%)
	Parents name	3 (3.95%)	3 (4.54%)
	Social Security number	3 (3.95%)	3 (4.54%)
	European Union Card	2 (2.63%)	2 (3.03%)
	Height	2 (2.63%)	0 (0.00%)
	Weight	1 (1.31%)	0 (0.00%)
	Parent's address	1 (1.31%)	0 (0.00%)
	Date of last medical check	1 (1.31%)	0 (0.00%)
Availability	Mobility	41 (53.95%)	41 (62.12%)
	Willingness to travel	41 (53.95%)	41 (62.12%)
Other data	Driving license	37 (48.68%)	37 (56.06%)

Table 4. Companies that increase potentially gender discriminatory items (worse policy)

28

Categories	Years	A	B	C	D	E	F	G	H	I	J	K
Sex / gender	2005	√	√	√	√							
	2009	√	√	√	√	√	√	√	√	√	√	√
Marital status	2005	√									√	√
	2009	√	√	√	√						√	√
Number of children	2005										√	
	2009	√									√	
Military service	2005											
	2009	√										

Note. $N = 11$. Each company is codified with a letter avoiding identification. These companies are different from Table 5.

Table 5. Companies that decrease potentially gender discriminatory items (better policy)

29

Categories	Years	R	S	T	U	V	W	X	Y	Z
Sex/gender	2005	√	√	√	√	√	√	√		
	2009	√	√	√		√		√		
Marital status	2005	√	√	√	√	√	√	√	√	
	2009	√								
Number of children	2005									√
	2009									
Military service	2005	√								
	2009									

Note. $N = 11$. Each company is codified with a letter avoiding identification. These companies are different from Table 4.

Table 6.Correlations between illegality scale, changes in policy, and annual returns

30

	1	2	3	4	5	6	7	8	9
1. Illegality scale (2005)	1								
2. Illegality scale (2009)	.80**	1							
3. Changes in illegality	.31**	-.32**	1						
4. Annual returns (2005)	-.10	-.25*	.23**	1					
5. Annual returns (2006)	-.16	-.19	.05	.59**	1				
6. Annual returns (2007)	-.14	-.06	-.31**	-.16	-.06	1			
7. Annual returns (2008)	.02	.12	-.15	-.22*	-.13	.37**	1		
8. Annual returns (2009)	-.01	-.00	-.01	.15	.09	-.09	-.27*	1	
9. Annual returns (Total)	-.15	-.12	-.04	.59**	.63**	.27*	.03	.66**	1

Note. $N = 66$. All the correlations are one-tailed.

* = p<.05; ** = p<.01.

31

Conclusions I

- **No discriminatory language found in discourses**
- Discriminatory items in application forms
  - Biased items against women in recruitment as part of the organizational culture
  - Need to respect equality in mass media (Act 3/2007. art. 39)
- Some relationships between annual returns and companies with discriminatory recruitment policy in application forms
- However, at a descriptive level, a rank-order the percentages of firms requesting various types of information suggests that those in the financial services and real estate and oil and energy categories request information most frequently that is potentially discriminatory and unfair, intrusive, and invasive of applicants' privacy. In contrast, *firms in the consumer goods category requested such information least frequently* both in 2005 and in 2009

32

Conclusions II

- Legislators and investigators need to collaborate to clarify:
  - art 5. 3/2007 about "difference of treatment based on a sex-related characteristic will not constitute discrimination in access to employment when, in light of the nature of the particular tasks concerned or the context in which they are performed, such a characteristic constitutes a **genuine and determining occupational requirement**, provided that the **objective is legitimate** and that the **requirement is proportionate**"
  - Act 15/99, about data processing in Internet, point out the importance of **adequate, accurate and non excessive personal information**. So, it is *urgent to clarify through applied research which data are adequate and relevant, and up to which point, regarding job access*

33

Tendencias para el futuro del reclutamiento

- Según Business 2 Community (2015):
  - Uso de teléfonos móviles para el reclutamiento;
  - Mayor uso de LinkedIn;
  - Incremento del intento de atraer candidatos qu eno están buscando empleo
  - Un mayor uso de Facebook y Twitter in recruitment, para ver sus perfiles en los SM, y uso de Facebook ya que se publicitan cada vez más anuncios de empleo en esta red social

Preguntas para la investigación (Davison, Maraist y Bing, 2011)

34

- Recruiting
  - 1. How does recruiting in this manner affect the applicant pool, particularly in terms of **quality?**
  - 2. Does use of these sites enable **organizations to target passive job-seekers?**
  - 3. Are **there differences in applicants** across different websites (e.g., MySpace, LinkedIn, Facebook, Twitter, etc.), given their potentially different clientele?
  - 4. Do applicants consider posting jobs on social networking sites more or less favorably, more or less fair, etc.?
  - 5. Are there generational differences in the acceptance of using social networking sites for recruiting?
  - 6. What are the applicant's reactions to an organization that has its own Facebook or Twitter presence? Does it attract a different type of applicant?

35

- **Screening and selection**
  - 1. Can personality be measured reliably and validly from web pages?
  - 2. Can other job-relevant characteristics, such as cognitive ability, person-organization fit, etc., be measured reliably and validly from web pages?
  - 3. How often do people provide inaccurate information on their profiles?
  - 4. What types of information are people providing inaccurately (e.g., jobs held, education)?
  - 5. How accurate is third-party provided information?
  - 6. What are the legal issues relevant to using social networking sites for screening and selection?

36

Muchas gracias por su atención
angarcia@uniovi.es

AVANCES EN MEDICIÓN ADAPTATIVA

Dr. Francisco J. Abad
Profesor Titular de la Facultad de Psicología
de la Universidad Autónoma de Madrid

Cátedra de Modelos y Aplicaciones Psicométricos
8º Seminario
24 de junio de 2016 | Facultad de Psicología de la UAM

Avances en Medición Adaptativa

Avances en medición adaptativa

Francisco José Abad García
Universidad Autónoma de Madrid

Avances en Medición Adaptativa

- Los TAIs
- Multi-etapa
- Multi-dimensionales
- Ipsativos

TAI-Matrices

Fluid Multistage Intelligence Test

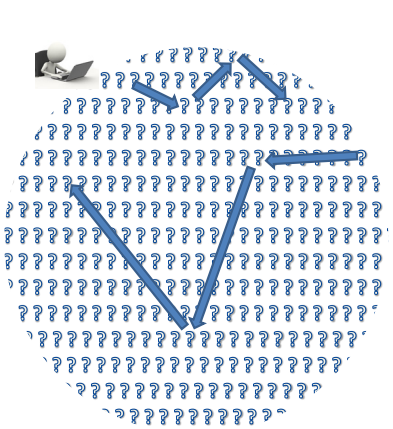
TAI-P

TAPAS

¿Qué es un TAI?

Avances en Medición Adaptativa

- Los TAIs
- Multi-etapa
- Multi-dimensionales
- Ipsativos



Banco de ítems

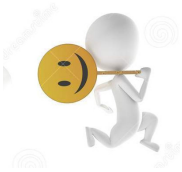
Algoritmo adaptativo

Ventajas de un TAI

Avances en Medición Adaptativa

Eficiencia y precisión homogénea.

Seguridad y "resistencia" al copie

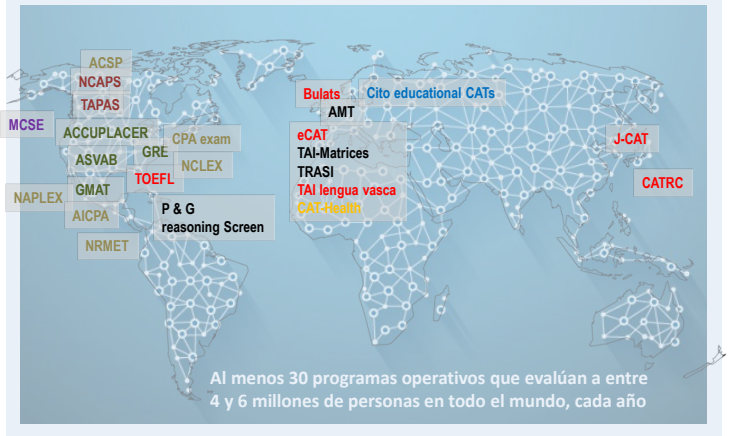


Motivacionales

VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

TAIs en el mundo

Avances en Medición Adaptativa

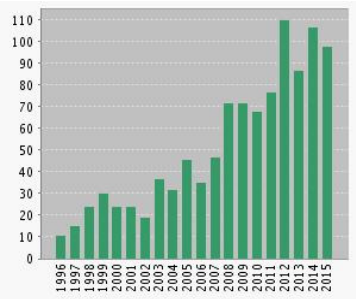


http://www.iacat.org/content/operational-cat-programs

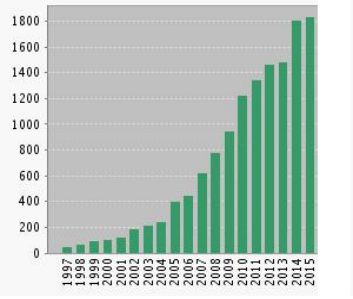
VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

Investigación TAIs

Avances en Medición Adaptativa



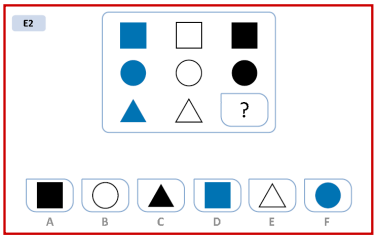
Nº de artículos publicados por año

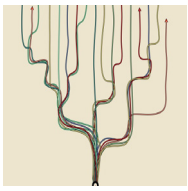


Nº de veces que se citan artículos sobre TAIs

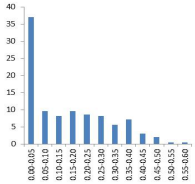
VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

TAI unidimensional
TAI-Matrices





Algoritmos de selección



Control de la exposición

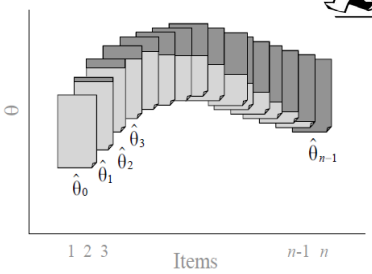
Banco final, 149 ítems ajustado a un modelo de Teoría de Respuesta al Ítem

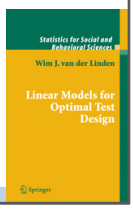
*Calibrado en una muestra de 12,280 personas (1,780 adultos)

TAI 27 ítems: rxx = .932 vs.
Fijo 36 ítems: rxx = .878

Control de restricciones

Tests "Sombra"

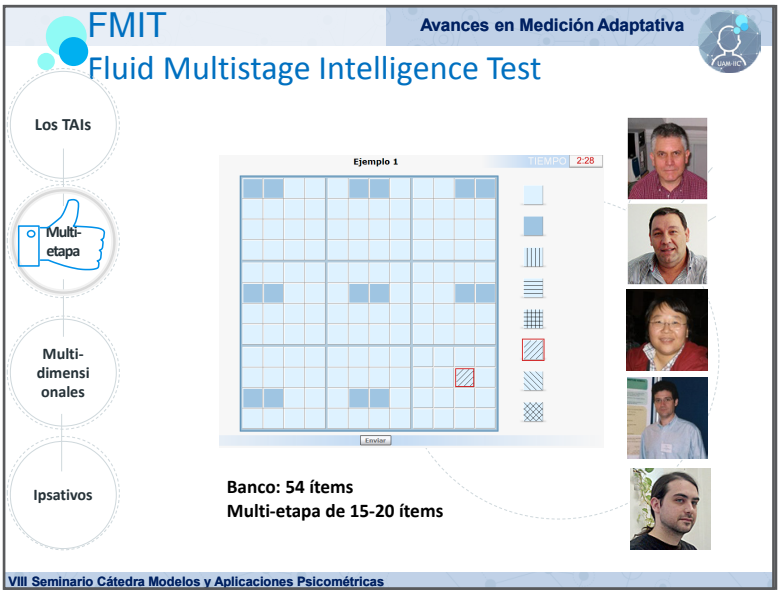


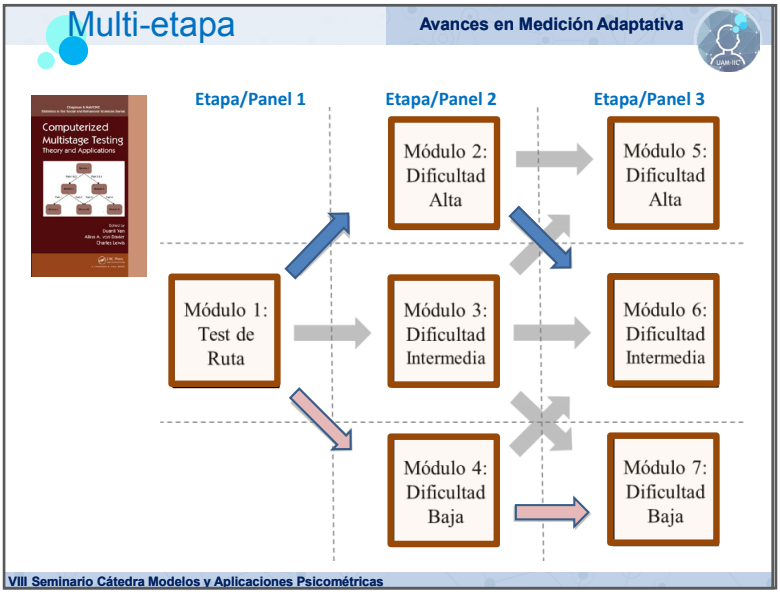


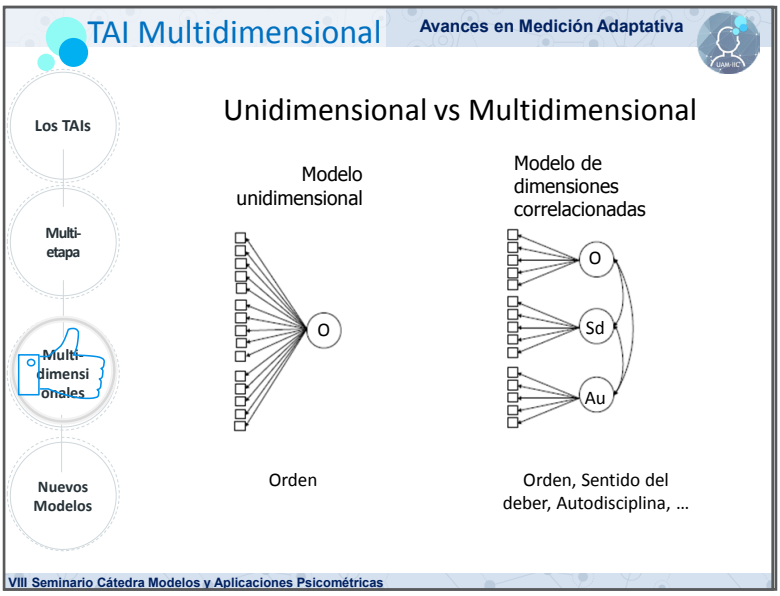
¿Problemas?

Diseño "sólo hacia delante". No se pueden dejar preguntas sin contestar

Demasiados test posibles


FMIT
Avances en Medición Adaptativa
Fluid Multistage Intelligence Test
Los TAIs
Multi-etapa
Multi-dimensionales
Ipsativos
Ejemplo 1
TIEMPO 2:28
Enviar
Banco: 54 ítems
Multi-etapa de 15-20 ítems
VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas


Multi-etapa
Avances en Medición Adaptativa
Computerized Multistage Testing
Theory and Applications
Etapa/Panel 1
Etapa/Panel 2
Etapa/Panel 3
Módulo 1: Test de Ruta
Módulo 2: Dificultad Alta
Módulo 3: Dificultad Intermedia
Módulo 4: Dificultad Baja
Módulo 5: Dificultad Alta
Módulo 6: Dificultad Intermedia
Módulo 7: Dificultad Baja
VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas


TAI Multidimensional
Avances en Medición Adaptativa
Los TAIs
Multi-etapa
Multi-dimensionales
Nuevos Modelos
Unidimensional vs Multidimensional
Modelo unidimensional
Modelo de dimensiones correlacionadas
O
O
Sd
Au
Orden
Orden, Sentido del deber, Autodisciplina, ...
VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

TAI Multidimensional

Avances en Medición Adaptativa

Los TAIs

Multi-etapa

Multi-dimensionales

Nuevos Modelos

Formas fijas → CATs 1.0 → CATs 3.0

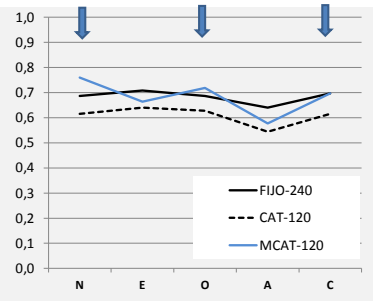
Table 1. Latent Correlations Between the Facets in the NEO PI-R

	N1	N2	N3	N4	N5	Average Correlation
N2	0.81					0.70
N3	0.91	0.85				
N4	0.87	0.72	0.87			
N5	0.40	0.59	0.41	0.34		
N6	0.85	0.79	0.87	0.82	0.38	
	E1	E2	E3	E4	E5	
E2	0.62					0.38
E3	0.30	0.32				
E4	0.40	0.35	0.46			
E5	0.18	0.33	0.38	0.26		
E6	0.59	0.45	0.33	0.48	0.23	
	O1	O2	O3	O4	O5	
O2	0.53					0.60
O3	0.68	0.63				
O4	0.75	0.55	0.55			
O5	0.72	0.64	0.75	0.69		
O6	0.49	0.47	0.67	0.36	0.48	
	A1	A2	A3	A4	A5	
A2	0.44					0.34
A3	0.45	0.42				
A4	0.30	0.42	0.37			
A5	0.19	0.47	0.29	0.36		
A6	0.23	0.25	0.34	0.22	0.31	
	C1	C2	C3	C4	C5	
C2	0.53					0.60
C3	0.68	0.63				
C4	0.75	0.55	0.55			
C5	0.72	0.64	0.75	0.69		
C6	0.49	0.47	0.67	0.36	0.48	

Avances en Medición Adaptativa

Un TAI para cada dominio de los Big Five

Fiabilidad



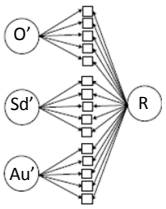
Adaptado de Makransky, 2012

TAI Multidimensional

Avances en Medición Adaptativa

Basados en un modelo bifactor

Modelo bi-factor



Orden, Sentido del deber, Autodisciplina, …

TAI Multidimensional Basados en un modelo bifactor

Avances en Medición Adaptativa

Banco de más de 300 ítems, cinco dimensiones, 6 facetas por dimensión.

Resultados preliminares para Responsabilidad (Facetas: Competencia, Orden, Sentido del deber, Necesidad de logro, Autodisciplina, Deliberación)
A partir de un banco de 54 ítems, se aplica un test de 12 ítems

Tipo de test	Correlaciones entre las puntuaciones obtenidas en el banco completo y las obtenidas según el tipo de test		Validez convergente (NEO-FFI)
	Dimensión general	facetas especificas (promedio)	
Test optimo	0.87		0.76
CAT unidimensional	0.88		0.78
CAT multidimensional		0.86	0.75
CAT bifactor	0.93	0.86	0.79
Banco completo	1	1	0.79

VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

TAIs Ipsativos

Avances en Medición Adaptativa

Los TAIs

Multi-etapa

Multi-dimensionales

Nuevos Modelos

Tailored Adaptive Personality Assessment System (TAPAS)

Drasgow y cols.,

1) TAI para medir 22 facetas de personalidad con entre 106 y 212 ítems por dimensión. 21 facetas se enmarcan dentro de los Cinco grandes:

- Responsabilidad: Laboriosidad, Organización, Auto-control, Tradicionalismo, Virtud, Responsabilidad social
- Estabilidad emocional: Ajuste , Calma, bienestar
- Afabilidad: Consideración, Altruismo, Cooperación
- Extraversión : Dominancia, Sociabilidad , Búsqueda de atención
- Apertura: Eficiencia intelectual , curiosidad , Ingenio , Estética , Tolerancia , Profundidad

2) Test "a medida" (tanto en el formato como en las facetas a medir).

3) Ejemplo: TAPAS-15D-CAT, 8 ítems por dimensión , 120 items: Tiempo medio 20 minutos,

VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

TAIs ipsativos

Avances en Medición Adaptativa

Likert vs. Elección-forzosa

Likert

Se le pide a una persona que diga en qué grado le representa un **ítem:**

Me llevo bien con los demás
Me suelo poner un estándar alto

Problema: En cada caso, la respuesta correcta es obvia.

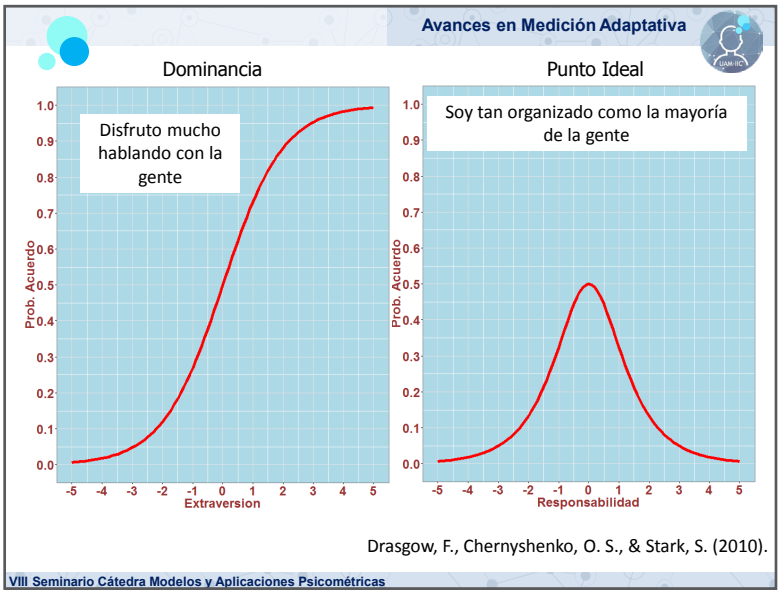
Elección-forzosa

Los **ítems** se agrupan en **bloques** igualados en deseabilidad social y se le pide a la persona que elija el que *más* le representa:

Me gusta decirle a la gente lo que debe hacer .

Devuelvo las llamadas telefónicas con prontitud.

VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas



TAIs ipsativos

Avances en Medición Adaptativa

Resistencia del formato al "falseamiento"

Comparación TAI vs test fijo

TAI vs fijo (10 dimensiones, 10 ítems por dimensión, 50 bloques, 5 unidimensionales) r = .90 vs. 84

Validez TAPAS 95-s (12 facetas)

Table 2
Incremental Validities of TAPAS-95s Scores for Predicting Selected Criteria

Criterion	N	AFQT only	AFQT + TAPAS	Δ *R*
Adjustment to army life	523	.13	.36	.23*
Army physical fitness test score	522	.04	.30	.26*
Number of disciplinary incidents	523	.11	.27	.17*
6-month attrition	1,694	.05	.24	.19*
Average technical training exam scores	585	.14	.23	.10*

Note. TAPAS = Tailored Adaptive Personality Assessment System; AFQT = Armed Forces Qualification Test.
* $p < .05$.

Tomado de Stark et al., 2014

VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

Otros temas

Avances en Medición Adaptativa

Mantenimiento del banco: detección del deterioro en los parámetros (Zhang, 2013; Zhang & Li, 2014) y calibración on-line para los nuevos ítems (Chang y Lu, 2010).

Pruebas de verificación (Egberink et al., 2010; Guo & Drasgow, 2010; Makransky & Glas, 2011).

Uso de información adicional: Uso de los tiempos de respuesta para mejorar la selección de los ítems o para detectar patrones extraños (van der Linden & van Krimpen-Stoop, 2003; van der Linden, 2008).

VIII Seminario Cátedra Modelos y Aplicaciones Psicométricas

Verificación…

Avances en Medición Adaptativa

Procedimiento verificación eCAT-grammar

TEST aplicado en condiciones UIT (30 items) → TEST Verificación (10 items) → ¿"Cheater"? → CONTINúA TEST Verificación (20 items más)

Procedimientos estadísticos:

Z Guo corregida

$$Z = \frac{\hat{\theta}_u - \hat{\theta}_v}{\sqrt{Se_u^2 + Se_v^2}}$$

	$\Delta\theta$	
θ_v	1	2
0	.83	.99
1	.80	

From the 954 examinees who were assessed in the proctored CAT, 132 (13.84%) were detected as cheaters

Referencias

Avances en Medición Adaptativa

TAIs multietapa:

- Yan, D., von Davier, A. A., & Lewis, C. (Eds.). (2016). *Computerized multistage testing: Theory and applications*. CRC Press.
- http://www.iic.uam.es/catedras/Test-Multietapa-inteligencia-fluidaR.pdf

TAIs-personalidad:

- Makransky, G., Mortensen, E. L., & Glas, C. A. (2012). Improving personality facet scores with multidimensional computer adaptive testing: an illustration with the NEO PI-R. *Assessment*.

TAIS-ipsativos:

- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right!. *Industrial and Organizational Psychology*, *3*(4), 465-476.
- Stark, S., Chernyshenko, O. S., Drasgow, F., & White, L. A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods*
- Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions.*Military Psychology*, *26*(3), 153-164.
- Usami, S., Sakamoto, A., Naito, J., & Abe, Y. (2016). Developing Pairwise Preference-Based Personality Test and Experimental Investigation of Its Resistance to Faking Effect by Item Response Model. *International Journal of Testing*, 1-21.

Referencias

Avances en Medición Adaptativa

Otros:

- Chang, Y. C. I., & Lu, H. Y. (2010). Online calibration via variable length computerized adaptive testing. *Psychometrika*, *75*(1), 140-157.
- Egberink, I. J., Meijer, R. R., Veldkamp, B. P., Schakel, L., & Smid, N. G. (2010). Detection of aberrant item score patterns in computerized adaptive testing: An empirical example using the CUSUM. *Personality and Individual Differences*, *48*(8), 921-925.
- Guo, J., & Drasgow, F. (2010). Identifying Cheating on Unproctored Internet Tests: The Z-test and the likelihood ratio test. *International Journal of Selection and Assessment*, *18*(4), 351-364.
- Makransky, G., & Glas, C. A. (2011). Unproctored Internet Test Verification Using Adaptive Confirmation Testing. *Organizational Research Methods*, *14*(4), 608-630.
- van der Linden, W. J., & van Krimpen-Stoop, E. M. (2003). Using response times to detect aberrant responses in computerized adaptive testing.*Psychometrika*, *68*(2), 251-265.
- van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics*, *33*(1), 5-20.
- Zhang, J. (2013). A sequential procedure for detecting compromised items in the item pool of a CAT system. *Applied Psychological Measurement*, 38, 87-104.
- Zhang, J., & Li, J. (2016). Monitoring Items in Real Time to Enhance CAT Security. *Journal of Educational Measurement*, *53*(2), 131-151.

RESUMEN POSTERS

Construcción de una herramienta de evaluación online para la medición de valores personales y organizacionales.

—

Virginia Arranz, Beatriz Lucía y Sonia Rodríguez.
Instituto de Ingeniería del Conocimiento

Las propuestas dirigidas a implementar modelos de dirección basados en valores están ocupando en la actualidad un lugar indiscutible en el entorno organizacional como complemento idóneo a los actuales modelos de gestión basados en competencias. De ahí el creciente interés por desarrollar herramientas de medición de valores en el trabajo que puedan dar respuesta a una doble demanda por parte de las organizaciones, por un lado para identificar de manera participativa cuáles son los valores corporativos y, por otro lado, para obtener una medida objetiva del impacto de las prácticas organizacionales derivadas de la implantación de modelos de gestión por valores.

El objetivo de este trabajo es presentar una herramienta de evaluación de valores personales y organizacionales, desarrollada por el Instituto de Ingeniería del Conocimiento (IIC) para lograr los anteriores objetivos. La prueba, de administración y corrección informatizada, está constituida por tres escalas que permiten obtener información de las siguientes tres variables: a) la percepción de ajuste y conflicto que el empleado tiene entre sus valores y los de su empresa b) la jerarquía de valores del empleado y la que percibe en su organización, así como la relación entre ambas y c) la coherencia personal y organizacional entre valores expuestos y valores reales o puestos en práctica. El modelo teórico que subyace a la prueba está basado en el Modelo Triaxial de valores emocionales, pragmáticos y éticos de la Dirección por Valores (DpV) propuesto por García, S. y Dolan, S. en 2009.

—

Correspondencia:
Virginia Arranz Monje
virginia.arranz@iic.uam.es

¿Tienes 10 minutos para predecir tu potencial comercial? - diseño de un SJQ para selección de Comerciales.

—

Katharina Lochner, Diego Martinez, Elena Grande, Beatriz Dosset, Patricia Murillo, Nora Nienhaus y Alexander Lemhoefer.
cut-e Spain y Securitas Direct

Las situaciones y el contexto importan. Tomar una decisión no es un acto en "vacío", sino el resultado de un complejo proceso de evaluación y juicio situacional. ¿Existen decisiones correctas o incorrectas, o más o menos apropiadas? ¿Puede un comportamiento ser adecuado en un momento y contraindicado en otro? Tal es la complejidad de la "realidad del día a día"; y nada más ejemplar que las grandes diferencias que hay en la función de comercial de una organización a otra. Este estudio presenta cómo evaluar ágil y eficazmente el ajuste de una persona a un puesto comercial específico.

Con el objetivo de potenciar su proceso de selección de comerciales, cut-e desarrolló para Securitas Direct un Cuestionario de Juicio Situacional (CJS), disponible en formato mobile y de menos de 10 minutos de duración (18 ítems, 25 en la versión inicial).

El proceso de construcción del CJS se basó en 25 horas de investigación de campo, incluyendo acompañamiento en la función de comercial.

El proceso de validación contó con una muestra de 93 personas, obteniendo resultados predictivamente positivos al contrastar las respuestas y resultados totales del CJS con el alcance de objetivos de ventas anuales: R=.731* y r= .595 (<.001).

—

Correspondencia:
Diego Martinez
diego.martinez@cut-e.com

Indicadores cuantitativos de calidad en procesos de selección.

—

David Martinez González, Jesús Martínez Soria y Belén Santero Guerra.
Grupo de Trabajo Psicología en la Evaluación de Personas en las Organizaciones del Colegio Oficial de Psicólogos de Madrid.

El presente estudio pretende desvelar cómo se utilizan indicadores cuantitativos de calidad en procesos de selección por parte de los departamentos de Recursos Humanos. Para ello, se realizó un análisis documental previo donde se identificaron más de 25 indicadores cuantitativos de calidad en procesos de selección.

Se realizaron entrevistas personales con responsables de selección de empresas españolas o con sede en España que tuvieran profesionalizado su proceso de selección. Las entrevistas se realizaban a partir de un cuestionario estructurado con más de 40 preguntas sobre la utilización de indicadores cuantitativos de gestión de procesos de selección.

Los resultados arrojan que aunque el 89% de las organizaciones consultadas emplean algún indicador de calidad en procesos de selección, el promedio de indicadores empleados tan solo llega al 43% de los indicadores analizados. También se observa que existen diferencias significativas en el uso de indicadores de calidad en las diferentes fases del proceso de selección. Como promedio, los indicadores más empleados arrojan información de valores absolutos, mientras que los menos empleados hacen referencia al coste de las diferentes fases del proceso de selección.

—

Correspondencia:
David Martínez González
dmartinez@actualgrupo.com

Evaluando Competencias Transversales a través de LinkedIn: un análisis exploratorio inicial.

—

Sonia Rodríguez, Ángel Álvarez, Francisco Álvarez y David Aguado.
Instituto de Ingeniería del Conocimiento y Grupo BLC

Cada vez más, la utilización de la información contenida en los perfiles de redes sociales de las personas es utilizada en los procesos de reclutamiento y selección desarrollados por las organizaciones. Los últimos datos al respecto de la utilización de estas redes sociales por los reclutadores ponen de manifiesto que, además de ser utilizadas con fines de reclutamiento también lo están siendo con fines de valoración de aspectos de los candidatos relacionados con su personalidad, actitudes o competencias. A pesar de que diversos estudios han puesto en relación las garantías de la información contenida en redes sociales no profesionales como FaceBook con la personalidad de los individuos, ningún estudio ha sido desarrollado con el objetivo de analizar el grado en el que a partir de la información contenida en una red social profesional como LinkedIn pueden inferirse las competencias transversales de los candidatos. En el estudio que se presenta se ofrecen los datos iniciales al comparar la información contenida en los perfiles de LinkedIn de 35 profesionales en búsqueda activa de empleo con sus puntuaciones obtenidas en un test de competencias transversales que mide las 8 grandes competencias del modelo de Bartram (2005). Los resultados muestran como algunas de las competencias están relacionadas con la forma en la que los candidatos ofrecen información y utilizan su perfil de LinkedIn.

—

Correspondencia:
David Aguado
david.aguado@iic.uam.es

Un nuevo test adaptativo informatizado para la evaluación de la capacidad intelectual general en el ámbito de los Recursos Humanos: el MATRICES-TAI.

—

Fernando Sánchez Sánchez[1], Pablo Santamaría[1] y Francisco José Abad[2].

[1]*Dpto. I+D+i de TEA Ediciones;* [2]*Universidad Autónoma de Madrid*

El Matrices-TAI es un test adaptativo informatizado (TAI) que mide el nivel de inteligencia o capacidad intelectual general. La aplicación se realiza por Internet y adapta la dificultad de los ítems presentados al nivel de aptitud de la persona, mostrando únicamente los más apropiados para estimar la aptitud general con la mayor precisión y en el menor tiempo posible.

El principal objetivo en el desarrollo del Matrices-TAI fue poder medir con precisión y eficiencia la mayoría de perfiles profesionales, desde los niveles más básicos a los más elevados. Se partió de la calibración (a partir de una muestra de más de 12.000 casos) de un banco final de 149 ítems. Se definieron las características del algoritmo adaptativo (estrategia de arranque, estimación del nivel de aptitud, selección de ítems, criterio de parada, control de la exposición…) y se decidieron los criterios óptimos de aplicación mediante simulación.

Los análisis realizados indican que las puntuaciones obtenidas con el algoritmo del Matrices-TAI poseen ausencia de sesgo y una precisión adecuada y homogénea en un amplio rango de niveles de aptitud, por lo que su uso en el ámbito de los recursos humanos resulta especialmente útil por su sencillez, brevedad y precisión de la medida.

—

Correspondencia:
Fernando Sánchez Sánchez
fernando.sanchez@teaediciones.com

Diseño y calibración de un nuevo banco de ítems basado en el modelo de los Cinco Grandes: aplicación del modelo bifactorial.

—

Francisco José Abad[1], María Dolores Nieto[1], Julio Olea[1], Alejandro Hernández[1], Luis Eduardo Garrido[2] y Juan Ramón Barrada[3].
[1]*Universidad Autónoma de Madrid,* [2]*Universidad Iberoamericana en Santo Domingo y* [3]*Universidad de Teruel*

La evaluación de la personalidad es una práctica habitual en contextos de selección de personal. Los cuestionarios empleados tradicionalmente para evaluar los Cinco Grandes (CG) presentan algunos inconvenientes (p.e., son demasiado extensos) que los Test Adaptativos Informatizados (TAIs) permiten solventar al evaluar con la misma precisión utilizando un número menor de ítems. En este trabajo se presenta el diseño y calibración de un nuevo banco de ítems de personalidad que constituirá la base substantiva de un TAI multidimensional (TAIM) para evaluar los CG. Para calibrar los ítems de cada una de las dimensiones de personalidad se empleó el modelo bifactorial (MB), el cual propone que cada ítem tiene pesos en un factor general (p.e., extraversión) y en un factor específico (p.e., gregarismo), ortogonales entre sí. El banco final está compuesto por 307 ítems con buenas propiedades psicométricas. Posteriormente, se realizó un estudio de simulación para comparar la eficiencia de diferentes TAIs (unidimensional, multidimensional tradicional y multidimensional basado en el MB) y la aplicación de una versión abreviada del test fijo para cada dimensión de personalidad. El TAI multidimensional basado en el MB demostró ser el método más eficiente, obteniendo estimaciones más precisas con menos de la mitad de los ítems.

—

Correspondencia:
María Dolores Nieto
mariadolores.nieto@estudiante.uam.es

Resistencia al falseamiento de las pruebas de elección forzosa: Efecto de la polaridad y de la dimensionalidad de los bloques de ítems.

—

Pedro Hontangas[1], Chloe Noya[1], Mercedes Merino[1] y Rodrigo Schames Kreitchmann[2].
[1]*Universidad de Valencia y* [2]*Universidad Autónoma de Madrid*

Se han realizado dos estudios para comprobar empíricamente el impacto del falseamiento en tests tipo Likert y de elección forzosa (EF). Ambos estudios utilizan pruebas de personalidad bajo el modelo de los Cinco Grandes, construidas a partir del IPIP (International Personality Item Pool): una tipo Likert (60 ítems) y otra de EF (30 bloques de dos ítems multidimensionales: 15 heteropolares y 15 homopolares).

En el primer estudio, 75 estudiantes de Psicología fueron asignados aleatoriamente a las condiciones de honestidad y falseamiento. Se les presentaron balanceadas las dos pruebas anteriores, más una nueva de EF con los 30 bloques homopolares. En el segundo estudio, 49 trabajadoras de una residencia de ancianos respondieron a los dos tests citados en primer lugar y a otro de EF con la misma cantidad de bloques uni y multidimensionales. La mitad de las trabajadoras recibieron instrucciones de responder honestamente y la otra mitad de falsear sus respuestas.

En general, los tests tipo Likert han presentado evidencias más claras de falseamiento que los de EF, aunque no en todas las dimensiones. En el poster se expondrán estos resultados y se discutirán los efectos de la polaridad y la dimensionalidad de los bloques.

—

Correspondencia:
Pedro Hontangas
hontanga@uv.es

Dos métodos de estimación para cuestionarios de elección forzosa por pares: MUPP-2PL frente a TRI Thurstoniana.

—

Daniel Morillo[1], Vicente Ponsoda[1], Francisco J. Abad[1], Iwin Leenen[2], Pedro Hontangas[3] y Jimmy de la Torre[4].

[1]*Facultad de Psicología (Universidad Autónoma de Madrid),* [2]*Instituto Nacional para la Evaluación de la Educación de México (México),* [3]Universidad de Valencia *y* [4]*The University of Hong Kong*

Los cuestionarios de elección forzosa (CEFs) están atrayendo la atención de los profesionales en selección de personal. Sin embargo, las puntuaciones ipsativas tradicionales suponen un problema para poder ordenar a los candidatos. La Teoría de Respuesta al Ítem (TRI) modela el proceso de respuesta, permitiendo la obtención de puntuaciones normativas.

En este trabajo se comparan los procedimientos de estimación recientemente propuestos para dos modelos: uno factorial confirmatorio (TRI Thurstoniana), y otro de estimación Bayesiana (MUPP-2PL). Para ello, se aplican los dos procedimientos a un CEF de 30 pares de elección forzosa, midiendo los Cinco Grandes, aplicado a una muestra de 567 estudiantes de grado universitario.

Ambos métodos producen resultados similares: los parámetros de los ítems y las personas tienen muy alta correlación por ambos métodos. Las estimaciones del MUPP-2PL para los ítems dan valores más realistas. Además, el MUPP-2PL estima con valores más aceptables algunos ítems que serían descartados según el método de TRI Thurstoniana. Las correlaciones entre escalas obtenidas con el MUPP-2PL son más moderadas, y se asemejan más a las obtenidas mediante un cuestionario de escala graduada. En conclusión, el MUPP-2PL parece funcionar mejor con el instrumento analizado.

—

Correspondencia:
Daniel Morillo Cuadrado
daniel.morillo@uam.es

Evaluación psicológica mediante juegos serios: Desarrollo y validación de medidas de impulsividad y auto-control.

—

Pei-Chun Shih Ma y Daniel Peña Molino.
Universidad Autónoma de Madrid y Apertia

Gomins es un videojuego de aventura espacial que además del entretemiento tiene como objetivo la evaluación de compentencias socioemocionales de los niños. El objetivo de este estudio es validar las medidas conductuales obtenidas en los juegos para la impulsividad y el autocontrol. 268 niños participaron en este estudio (178 niños, 86 niñas; rango de edad: 4 a 12, M = 8.35, DT = 2.06). Además de los juegos, los participantes fueron evaluados por psicólogos clínicos expertos en: control inhibitorio, planificación, perseverancia y demora del refuerzo. La correlación entre ambas medidas conductuales fue baja (r = -.180, p = .004). Las correlaciones entre índice de impulsividad y las variables clínicas fueron medias: control inhibitorio (r = -.269, p = .000), planificación (r = -.254, p = .000), perseverancia (r = -.244, p = .000), demora del refuerzo (r = -.222, p = .001). En cuanto al índice de autocontrol, sólo se encontró una correlación media con la demora del refuerzo (r = .214, p = .001). Es importante señalar que los patrones de correlaciones difieren según los rangos de edad (4-6, 7-8, 9-12). Los resultados de este estudio aportan cierta evidencia de validez concurrente a las medidas obtenidas en los juegos.

—

Correspondencia:
Pei-Chun Shih Ma
pei_chun.shih@uam.es

Validez y fiabilidad de las puntuaciones en Tests de Juicio Situacional: un nuevo enfoque basado en Modelos de Diagnóstico Cognitivo.

—

Miguel A. Sorrel[1], Julio Olea[1], Francisco J. Abad[1], Jimmy de la Torre[2], David Aguado[3] y Filip Lievens[4].

[1]*Universidad Autónoma de Madrid,* [2]*The University of Hong Kong,* [3]*Instituto de Ingeniera del Conocimiento y* [4]*Ghent University*

Los métodos tradicionales para evaluar la validez y fiabilidad de las puntuaciones han demostrado ser ineficaces en el caso de los Test de Juicio Situacional (TJS). Por ejemplo, las técnicas de análisis factorial suelen producir soluciones sin sentido y los supuestos del coeficiente alfa de Cronbach no pueden ser asumidos debido a su compleja naturaleza multidimensional. En el presente trabajo describimos cómo los Modelos de Diagnóstico Cognitivo (MDC) pueden ser la base de un nuevo enfoque que permita superar estas limitaciones y ofrecer ventajas adicionales para puntuar y entender mejor los TJS. Ilustraremos nuestra propuesta empleando datos recogidos para un TJS de 23 ítems que explora competencias del ámbito educativo (p. ej., hábitos de estudio, actitudes hacia la educación). Las clasificaciones obtenidas fueron fiables y estuvieron significativamente relacionadas con variables teóricamente relevantes como es el caso de la nota media en la carrera. Los resultados muestran que los MDC son herramientas útiles para puntuar test que son multidimensionales también a nivel de ítem, como es el caso de los TJS. En conclusión, encontramos que los MDC pueden ayudarnos a avanzar en una de las principales líneas de investigación en el ámbito de los TJS: la exploración de los constructos evaluados.

—

Correspondencia:
Miguel A. Sorrel
miguel.sorrel@uam.es

CÁTEDRA DE
MODELOS Y APLICACIONES PSICOMÉTRICOS

UAM-IIC

http://www.iic.uam.es/catedras/map