

Interpreting scores

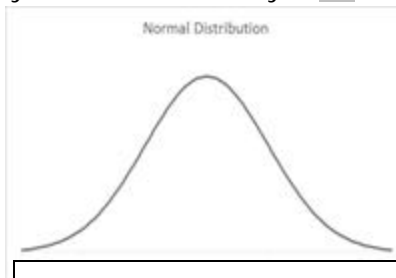
- ▶ A test provides scores on arbitrary scale
 - ▶ Is 35 correct good?
 - ▶ Is 67% correct poor?
- ▶ There are two versions of a test
 - ▶ Are scores on the two versions equivalent?
- ▶ Standardisation helps address these issues

Standardised scores

- ▶ Provide a number scale with known meaning
- ▶ Help interpret scores by
 - ▶ Comparing to a relevant group (norm referenced)
 - ▶ Comparing to a given standard (criterion referenced)
- ▶ Criterion referencing common in education
- ▶ Norm referencing more usual in employment contexts
 - ▶ For selection want the best available candidate – not a specific standard

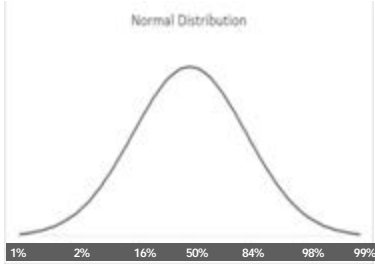
Scores typically distribute normally

- ▶ Most scores are in the middle of the range
- ▶ The more extreme the score the more infrequent
- ▶ I.e. there is a typical preference or level of performance but there are always a few people who are substantially different



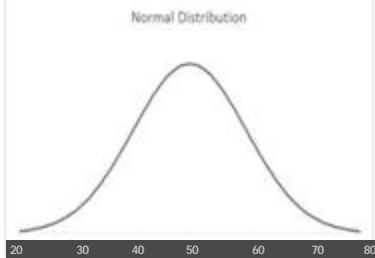
Percentile Scale

Percentage who score at that level or lower
67% score above the 33rd percentile
1% score above the 99th percentile
Percentiles spread scores in the middle or the range and collapse scores at the extremes
Widely used for ability measures and sometimes used with questionnaires
Score depends on comparison group



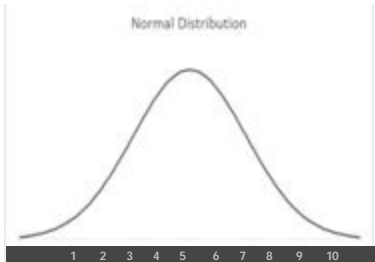
T Score

One of a number of standardised scores
Provides more separation at extremes and less at centre
Most differentiated scale in common use - typical used with ability measures
Can combine standardised scores
Average T score is 50
67% score between 40 and 60



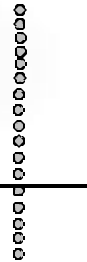
Sten Score

Less differentiated standardised score - 10 points
Provides more separation at extremes and less at centre
Typically used with personality scales
Can combine standardised scores
Average Sten score is 5.5 (not a realisable score)
67% score between 4 and 7
No middle point



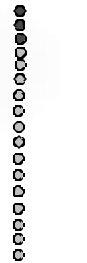
Using scores: Cut scores

- ▶ Performance higher than minimally acceptable level
 - ▶ E.g. accept scores 30th percentile and above
- ▶ Good for shortlisting
- ▶ Avoid setting cut score too high
 - ▶ Easy way to reduce numbers
 - ▶ Reduces selectivity on other factors
 - ▶ Can result in unfair discrimination



Using Scores: Top down selection

- ▶ Selecting the best from the pool
- ▶ Maximises impact of using test
- ▶ Restricts selectivity on other factors
- ▶ Can result in unfair discrimination



Using Tests in Selection

Designing a selection system

- ▶ Range of exercises - to measure as much of the job as possible
- ▶ Clear links to role and competencies
- ▶ Broad coverage of role
- ▶ Focus on elements that are less open to development

Using psychometric measures

- ▶ When is it appropriate to use tests and questionnaires?
- ▶ What measures to use?
- ▶ Where should they be placed in the selection process?
- ▶ How is the information used?

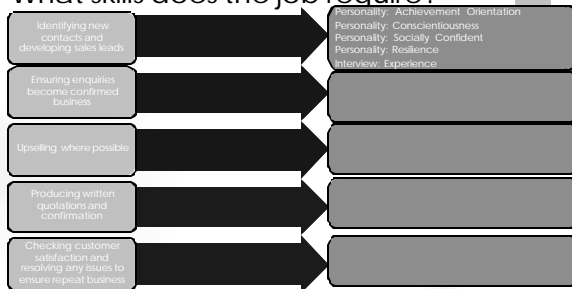
Example job: Room sales executive

- ▶ Main aim of role:
 - ▶ Increase the number of room sales in the hotel
- ▶ Reports to: Room sales manager
- ▶ Key responsibilities
 - ▶ Identifying new contacts and developing sales leads
 - ▶ Ensuring enquiries become confirmed business
 - ▶ Upselling where possible
 - ▶ Producing written quotations and confirmation
 - ▶ Checking customer satisfaction and resolving any issues to ensure repeat business

Developing Criteria

- ▶ Start with understanding the job
- ▶ Formal Job Analysis
 - ▶ Observations
 - ▶ Questionnaires
 - ▶ Critical Incident Technique
 - ▶ Repertory Grid
- ▶ For selection the focus is on elements that differentiate levels of performance

What skills does the job require?



Summary of Suggested Tools to use

- ▶ Personality Questionnaire
- ▶ Verbal Reasoning Test
- ▶ Numerical Reasoning Test
- ▶ Interview
- ▶ Written exercise
- ▶ Role play

Example Assessment Grid	PQ	Interview	Role Play	Written Exercise	Verbal Ability Test
Impact & Influence	*		**	**	
Initiative	*	**		**	
Building Collaborative Relationships	*	**	**	*	
Self-Awareness & Resilience	*	**			
Problem solving	*			**	**

Selecting Ability Tests

- ▶ Measures something important for the role
- ▶ At the correct level
- ▶ Available in local language
- ▶ Has relevant norm group - or develop own
- ▶ With suitable content
 - ▶ Accessible
 - ▶ Appropriate
- ▶ Reliable
- ▶ Valid

Using Ability Tests

- ▶ Requires qualified test user
- ▶ Where in the process
 - ▶ Early: Sifting out unsuitable candidates
 - ▶ Late: Matching against person specification

Administration

- ▶ Supervised
 - ▶ Control identity, cheating
 - ▶ Can be expensive, time consuming
- ▶ Remote
 - ▶ Random generated online tests
 - ▶ Convenient
 - ▶ Supports early sifting
 - ▶ Open to abuse
 - ▶ Need to re-test later

Interpreting results

- ▶ Ensure there is a suitable norm group
 - ▶ Similar educational level
 - ▶ Relevant background and experience
 - ▶ Good sampling
- ▶ Wrong norm group
 - ▶ Tends to bunch scores
 - ▶ Misleads regarding candidate abilities
 - ▶ Can create unfairness

Level of performance

- ▶ Consider job requirements
 - ▶ How complex is need relative to test
 - ▶ How important is need in context of whole role
 - ▶ Is the capacity used occasionally or constantly?
 - ▶ Is it relevant to critical elements of the role?
 - ▶ Is it only required during training and familiarisation or ongoing need?

Score

Test Score



The candidate scored seven on the verbal reasoning test which is an above average level compared to the comparison group. The score falls at the 77th percentile so the candidate scored better than 77% of the comparison group.

Speed of working

The candidate's pace of work was typical of those in the comparison group falling at seven.

Caveats to score

- ▶ How long is the score valid
- ▶ What threats to validity are there
 - ▶ E.g. test was completed without supervision
- ▶ How is it appropriate to use the score

Selecting personality questionnaires

- ▶ Model sufficiently detailed for need
 - ▶ Big 5 for quick review
 - ▶ More detailed model for most uses
- ▶ Language and complexity appropriate for group
- ▶ Relevant norm group
- ▶ Appropriate reports available

Language availability

- ▶ There are a number of questionnaires available in a broad range of languages
- ▶ OPG32 (includes Estonian)
- ▶ WAVE
- ▶ 16PF
- ▶ PAPI
- ▶ MBTI
- ▶ Dimensions
- ▶ NEO PI-R
- ▶ HPI

Administration

- ▶ Supervised
 - ▶ Control identity
 - ▶ Some forms of cheating not controlled
 - ▶ Can be expensive, time consuming
- ▶ Remote
 - ▶ Usually single version of questionnaire
 - ▶ Convenient
 - ▶ Supports early sifting
 - ▶ Does not require re-test unless indication of unreliability in results

Personality Questionnaire Results

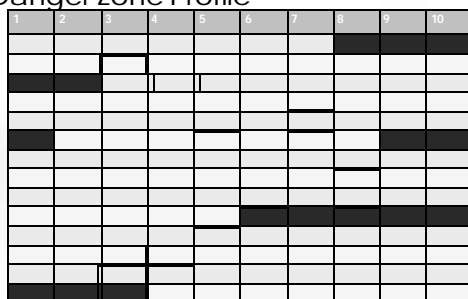
- ▶ More complex than ability measures because multiple scales
- ▶ Generally reported on less differentiated score scales
- ▶ Trained test users can receive profiles of scores
- ▶ Untrained test users receive narrative reports without scales or scores
- ▶ Reports mostly computer generated although can be hand written

Options for reporting

- ▶ Based on questionnaire scales
 - ▶ Description of individual's style
 - ▶ Reader needs to relate this to role requirements
- ▶ Using competency framework
 - ▶ More work relevant. Report writer maps scales to competencies
- ▶ Using a typing model
 - ▶ E.g. team types, leadership style

RELATIONSHIPS WITH PEOPLE		1	2	3	4	5	6	7	8	9	10	
3	Less comfortable persuading others and influencing their opinions; prefers not to engage in selling.	Influence										Enjoys influencing others, selling ideas and negotiating.
5	Finds it difficult to hide irritation or annoyance. Can say things that will upset others.	Self-Restraint										Controls negative emotions. Avoids showing irritation and can remain patient in face of some provocation
5	Not particularly interested in the needs of others. Can be insensitive to what others want.	Agreeable										Likes to help others and will be sensitive to others needs and problems
8	Enjoys talking about own achievements and likes to receive praise	Modest										Prefers not talk about self and own achievements. Can feel uncomfortable and prated by others.
6	Works best on own. Prefers not to have to work closely with others in a team	Team Player										Enjoys work with others in a team and dislikes working alone for long periods
5	Can lack social confidence. Shy with strangers	Social Confidence										Lively and sociable. Talkative and outgoing in a group -- even when they are strangers

Danger Zone Profile



Competency based report

Competency: Teamwork					
✓ Generally willing to help others	Competency Score				
✗ May not always finish own tasks	1	2	3	4	5
✗ May not recognise when others are upset					
Competency: Coping with pressure					
✓ Tends not to worry about problems	Competency Score				
✓ Has an optimistic approach	1	2	3	4	5
✓ Willing to try again after a failure					

Text based competency report

Sophie is likely to be as thorough and as organised as most other people and will have some concern for the quality and the detail of her work, but without being a perfectionist. Probably reasonably attentive to detail, Sophie may need to make more effort than most to sustain a consistent level of high performance.

Using personality data

- ▶ Danger zone to shortlist
- ▶ Generate competency scores
- ▶ Generate suitability scores
- ▶ Generate issues to explore at interview
- ▶ Qualitative integration of information with other areas

Issues in Test Use

User Training

- ▶ Technical knowledge and skills are required to use tests effectively
- ▶ Tests are powerful and there is the potential for negative impact if not used appropriately
- ▶ Reputable test publishers will only provide tests to people with appropriate training
- ▶ Euro Test User Certificate operated by EFPA is one type of qualification.

Candidate experience

- ▶ People typically take tests when they are going through life changing events e.g. applying for a job
 - ▶ High stakes situation
- ▶ Important that assessment treats candidates in a respectful manner
- ▶ Try to make the process as comfortable as possible for the candidate
- ▶ Candidates should get something out of the assessment process even if they do not get the job

Communication with candidates

- ▶ Explain the process in advance
- ▶ Provide clear information about what candidates will be required to do
 - ▶ Helps to deal with candidate anxiety
 - ▶ Candidate should understand if and how they should prepare
 - ▶ Provide access to practice materials where appropriate
- ▶ Explain how the information will be used and who will have access to it

On the day

- ▶ Ensure that all those who come into contact with candidates treat people with respect
- ▶ Allow for candidate nerves
 - ▶ Use a business-like but friendly manner
 - ▶ Don't get impatient if you have to explain things again
- ▶ Explain again
 - ▶ what is being assessed
 - ▶ what will happen to the results
 - ▶ how feedback will be provided

Provide feedback

- ▶ Candidates feel uncomfortable because they do not know what information they have provided
- ▶ Feedback encourages honesty
- ▶ Gives something to candidate for their time in going through the process

Language of testing - Ability

- ▶ People perform better in their primary language
- ▶ Can confuse ability and language skills
 - ▶ Better to measure separately
- ▶ If testing in another language – consider the impact of language
- ▶ Consider using non-verbal tests if appropriate
- ▶ Norms may not be appropriate for test takers with different language skills

Language of Testing - Personality

- ▶ Personality questionnaire results remarkably similar across languages
- ▶ But measurement may be impaired with non-primary language testing
- ▶ May be appropriate to help with vocabulary
- ▶ Some evidence that language of completion has some impact on profiles

Data Protection

- ▶ Confidentiality for candidate
- ▶ Storage of materials
- ▶ Permission to use scores
- ▶ Don't use scores without/beyond permission

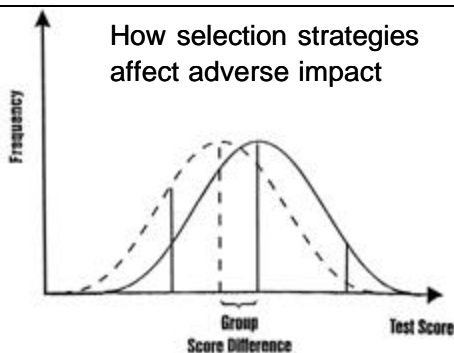
Fairness

- ▶ Tests promote fairness because they are not affected by personal biases
- ▶ Well designed tests work across many groups
- ▶ When tests reflect job requirements their use will tend to promote fair selection
- ▶ Group differences are not necessarily bias

Unfairness

- ▶ Many tests do show group differences - so a less relevant tests could be unfair
- ▶ Measuring constructs that some groups have had less opportunity to develop
- ▶ Tests requiring higher levels of ability than required on the job
- ▶ Tests requiring irrelevant skills
 - ▶ Language
 - ▶ Reading speed
 - ▶ Content knowledge
- ▶ Using top down selection or high cuts scores when there are group differences

How selection strategies affect adverse impact



Disability Accommodation in Occupational Assessment

- ▶ Focus on assessing competence - not disability
- ▶ Identify barriers to demonstrating competence in assessment design
 - ▶ E.g. accessing information
 - ▶ Working with response methodology
 - ▶ Appropriateness of time constraints
- ▶ Adapt the assessment process to provide information equivalent to that for someone without a disability
 - ▶ E.g. Alternate information provision
 - ▶ Alternate response modality
 - ▶ Extended time allowance
 - ▶ Alternate location, administration conditions

Altering assessments

- ▶ Requires expertise in both disability and assessment
- ▶ Aim is to conserve construct validity
 - ▶ Why is this assessment being used in this context?
- ▶ Consider how the person might do this at the in the job
- ▶ Job accommodations won't always be appropriate for assessment
 - ▶ E.g. if training is required to use equipment provided

Process considerations

- ▶ Candidates may prefer not to disclose disability if not necessary
- ▶ Provide information so candidates can determine if they need to disclose disability
- ▶ Be clear:
 - ▶ That you are open to providing accommodations
 - ▶ How to request accommodations
 - ▶ Who to contact
- ▶ Be ready to arrange accommodations within the assessment timescales
 - ▶ But don't need to pre-prepare for every eventuality

How long are results valid

- ▶ Typical test results considered valid for up to 2 years
- ▶ Life events can lead to changes in longer periods – but most people have very stable results
- ▶ Never use tests for purpose that was not originally agreed without seeking further permission
 - ▶ Unlikely to be appropriate to use tests for redundancy decisions

Coaching

- ▶ Working with candidates to improve test scores
- ▶ Studies show impacts all types of selection not just tests
 - ▶ Interview performance strongly impacted by coaching
- ▶ Coaching to develop skill measured
 - ▶ Increase in score is valid
- ▶ Coaching to increase test scores
 - ▶ Pre-coaching test score more valid
 - ▶ Good tests less open to coaching
 - ▶ Providing practice materials lessens the impact of coaching

Personality: Can we trust responses?

- ▶ Faking or social desirability
- ▶ Self Deception / Low self insight
- ▶ Need to Please
- ▶ High Self Esteem
- ▶ High Conformity
- ▶ Impression Management

Social desirability scales

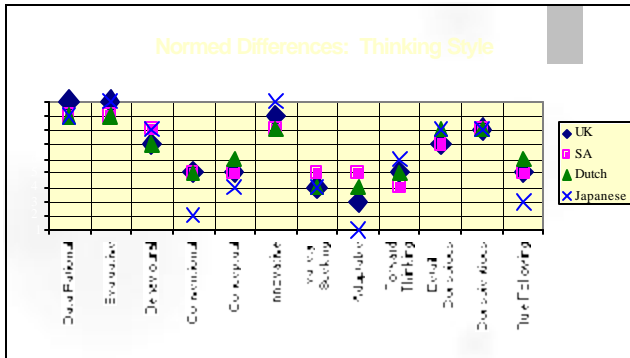
- ▶ Unlikely virtues
 - ▶ I always give people my full attention
 - ▶ There are times when I have been unfair to other people
 - ▶ I don't talk about other people when they are not present
- ▶ Desirable traits
 - ▶ I am reliable
 - ▶ I practice what I preach

Social Desirability Findings

- ▶ When instructed to fake people can change scores greatly (SD+)
- ▶ Applicants score around 0.5 SD higher than incumbents (Hough 1998)
- ▶ Warnings not to distort reduce differences
- ▶ Use of rankings helps control social desirability responding
- ▶ Validity findings strong but marginally weaker with applicants than incumbents
- ▶ Social desirability scales can have own validity but do not improve prediction as moderators

Cultural Equivalence

- ▶ Surprising cross cultural similarity in findings
- ▶ Need to develop translatable models and items
- ▶ Group differences tend to be small
- ▶ Gender and age differences can be larger



Interpreting Cross Cultural Profiles

- ▶ Be aware of possible differences
 - ▶ people can learn to adapt behaviour
 - ▶ can't change dispositions
- ▶ Response styles can differ
 - ▶ e.g. higher need to please in Far East
- ▶ Different scales may be seen as 'socially desirable'
