Saville
assessment

International Accreditation Ability

Course Workbook

TRAINING

International Accreditation

CONTENTS

Introduction: SavilleAssessment

Psychometrics

- 2004: Founded by Professor Peter Saville
- 2005: Wave launched
- 2007: Swift combination tests launched
- 2013: Situational judgment tests launched
- 2015: New tests, new technology
- 2015-2017: Saville Assessment, A Willis Towers Watson Company
- 2017: Leadership Impact and Risk launches
- 2019: Match 6.5 launched
- 2021: Swift Global launched
- 2022: Wave-i launched
- 2023: Saville Assessment acquired by Tenzing private equity firm

Talent Assessment Solutions

Hire Talent

Improve Quality of Hire

Pinpoint the drivers for success, identify the right people for the right roles and maximize talent acquisition metrics.

Build Talent

Maximize Talent Effectiveness

Identify potential, develop performance, create agile teams and improve workplace productivity.

Lead Talent

Transform Leadership Effectiveness

Identify, select and develop leaders who will create the most positive impact on your organization and accelerate exceptional results.

Objectives

By the end of the course, you will be able to:

Ability

- choose assessments
- administer tests
- interpret test scores
- feedback test results
- apply tests fairly

Notes

About Saville Assessment

The journey of Saville Assessment started in 2004 when a team of assessment specialists came together. The team comprised experts in Occupational Psychology, Business Consulting and Information Technology, with the goal of transforming and revolutionizing assessment around the world.

Our assessment tools are available in over 40 languages; please contact us for more information.

From 2015 to 2023 Saville Assessment was acquired by Towers Watson.'global organization.

In 2023 Saville Asssessment was acquired by Tenzing, a private equity firm.

A Brief History

2004 – Saville Consulting is founded

'Assessment Guru' Professor Peter Saville recruited a team of assessment experts/psychometricians to deliver his vision of transforming assessment around the world.

2005 – Wave

A new era of personality questionnaires arrives, offering the highest validity on the market and the deepest insight into an individual's motives, talents and workplace potential.

2007 – Swift combination ability tests

Faster, smarter ability testing boasting a fresh, modern look and feel, and the only portfolio to include combination tests measuring several sub-areas in one assessment.

2009 – Item-banked ability tests

Introduction of item-banks across our ability test portfolio to ensure greater security in online assessment.

2013 – Situational Judgment Tests

Custom, multi-media SJTs combining psychometric expertise with the latest technology breaks boundaries with a fast, engaging, powerfully branded volume assessment tool.

2015 – New tests, new technology

The first psychometric test publisher to have tablet-administered assessments and lead the way with utilizing technology.

2015 - 2017 – Saville Assessment, A Willis Towers Watson Company

Became the talent assessment part of the leading global advisory, broking and solutions company, helping clients around the world turn risk into a path for growth.

2017: Leadership Impact and Risk launches

Bridging the gap between behavioral skills potential and leadership impact to support with leadership recruitment and development.

2019: Match 6.5 launched

A new behavior questionnaire which uses the power and validity of Wave to understand a candidate's suitability for a role in just 6.5 minutes.

2023: Saville Assessment Aquired by Tenzing

Tenzing is a private equity firm that invests in high growth businesses.

Module 1: Introduction to Testing

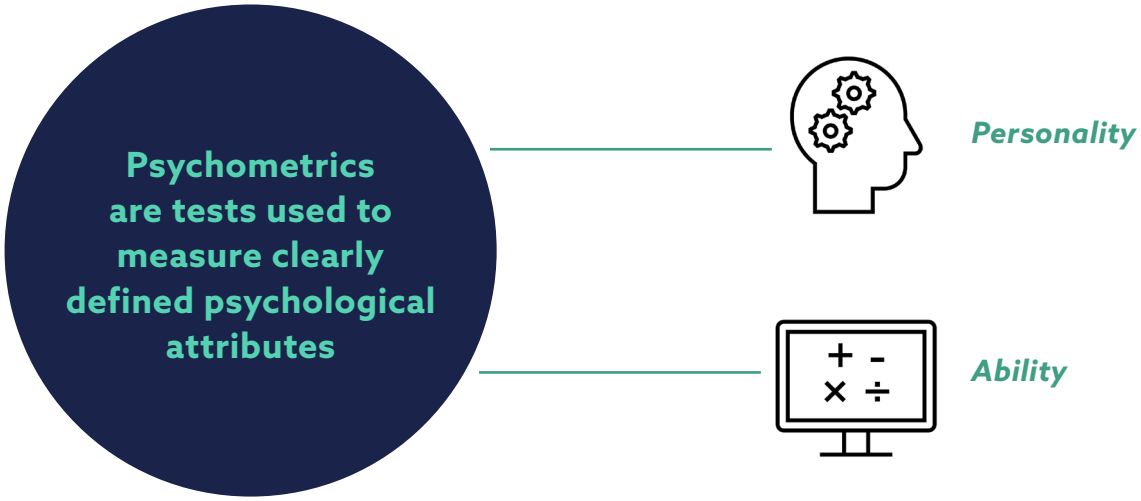
Projective Tests – Inkblot Test

Notes

Projective Tests – Thematic Apperception Test

Notes

What is a Psychometric Test?



An assessment of a psychological attribute, typically scored using a numerical scale or category system, to describe individual differences.

'Will Do' Assessments of Typical Performance

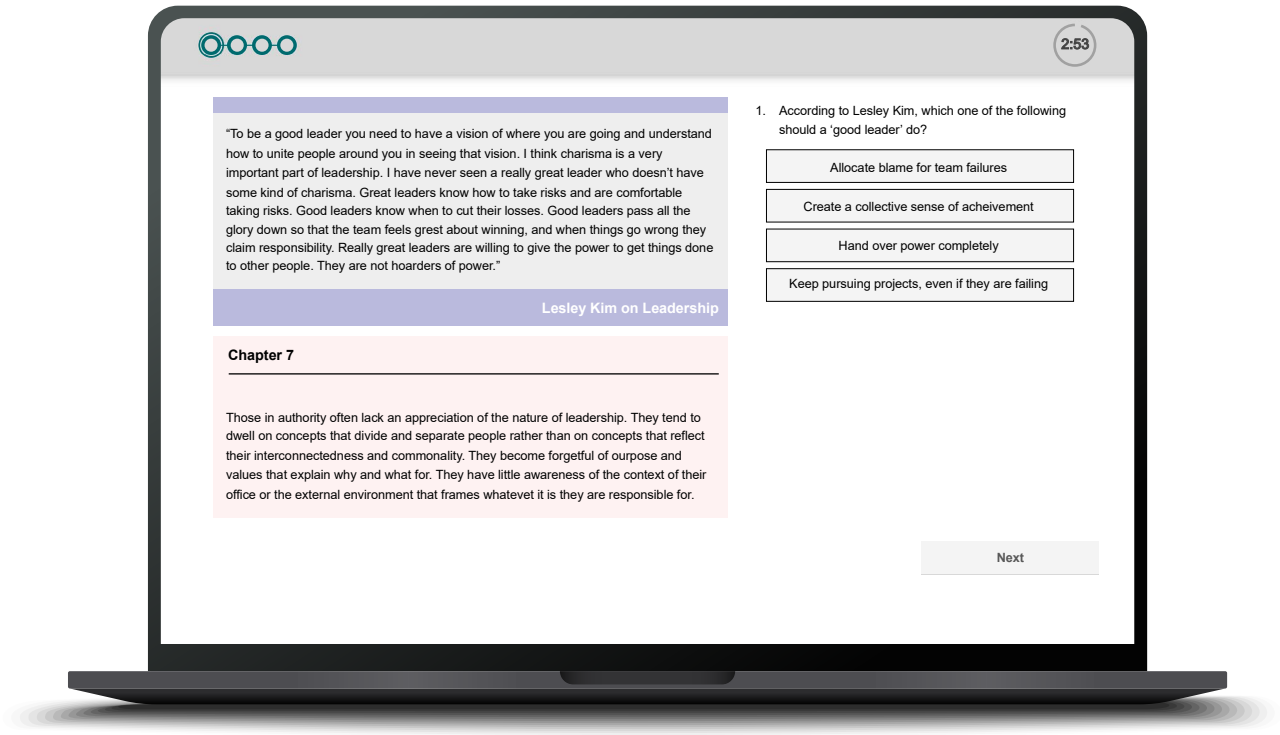
- Include self-report questionnaires without time limits
- 'Right' and 'wrong' can vary depending on context

Wave Professional Styles example:

‘Can Do’ Ability Tests of Maximum Performance

- Include ability tests of ability, IQ and attainment:
  - Ability: predict what someone will be able to learn or do in the future, e.g. Saville tests
  - IQ: current level of intellect/cognitive ability, e.g. Wechsler Adult Intelligence Scale
  - Attainment: measure current level of knowledge understanding or skill, e.g. driving test
- Often with strict time limits
- Clear right or wrong answers

Verbal analysis example:



Key Facts: Ability Tests

- Benchmarks against external group
- Have good validity so are a strong predictor of performance
- Measures lots of different types of ability
- Efficient online assessment
- Fair and consistent treatment of candidates
- Supplements other sources of information
- Sophisticated question banking to counter cheating

Notes:

'Can-do and Will-do Assessments

Will-do Tests

These measure typical performance, examples of which are listed below:

- Interest inventories/questionnaires
- Personality questionnaires
- Motivation questionnaires
- Job performance
- Attitude surveys
- 360 degree assessments

Interest inventories/questionnaires measure the things an individual is interested in. This type of information may be useful in career guidance. Personality questionnaires look at styles of behavior, for example the Occupational Personality Questionnaire (Saville et al, 1984) and the Professional Styles and Focus Styles versions of Saville Assessment's Wave. Motivation questionnaires measure what people want to do. Note: this can also be measured by the Wave questionnaire detailed above. Rating scales look at measures of job performance. Attitude surveys are often of great interest in market research. 360 degree assessments ask for ratings from bosses, colleagues and subordinates. Saville Assessment has developed the Wave Performance 360 questionnaire to gather self and other ratings online.

Can-do Tests

These assess maximum candidate performance, examples of which are listed below:

- Aptitude
- Achievement/attainment
- Intelligence tests (IQ)
- In-tray
- Work sample
- Trainability tests

Aptitude tests measure abilities that underpin future potential – examples include Saville Assessment's verbal, numerical and diagrammatic analysis tests. Achievement/attainment tests look at an individual's level of current knowledge – examples include school exams or a driving theory test. Intelligence tests (IQ) are a mixture of aptitude and attainment, one common measure of IQ is the Wechsler Adult Intelligence Scale. In-tray exercises/business simulation exercises are tests which assess skills at particular tasks and are often very useful in assessment centers. Work sample tests present applicants for a job with a sample of the work they will be expected to undertake in the job. Trainability tests assess how well individuals respond to training.

Key Benefits: Aptitude Tests

This Swift Occupational Ability course focuses on the Can-do – Maximum Performance assessments. These are some of the key benefits of using ability tests and we will expand on them throughout the course:

- Benchmarks against external group
- Single most valid predictor of work performance
- Measures lits of different types of ability
- Efficient online assessment
- Fair and consistent treatment of candidates
- Supplements other sources of information
- Sophisticated question banking to protect the security of the content

Module 2: Job Analysis

Job Analysis

- Job analysis is a multi-method approach that is used for different purposes including:
  - Defining role profiles/job descriptions/person specifications
  - Job sizing for pay grading
  - Developing a framework of criteria for assessment e.g. skills potential
- In assessment, good job analysis focuses on things that can be defined clearly and measured well

Common Methods of Job Analysis

- Structured interviews
  - Job holders e.g. critical incident identification
  - Line managers e.g. repertory grid comparisons
- Job content reviews
  - Diaries
  - Observing the job
  - Doing the job
  - Task/job analysis questionnaires
- Validation research

Notes:

Job Analysis

An important concept of Job Analysis is that the analysis is conducted on the job, not the person. While data may be collected from incumbents through interviews or questionnaires, the product of the analysis is a description or specification of the job, not a description of the person to be hired. Job Analysis is an essential pre-requisite to choosing which psychometric tests and questionnaires to use. In assessment, good job analysis focuses on things that can be defined clearly and measured well.

What is Job Analysis?

Job Analysis is a detailed process to identify and determine the particular job duties and requirements, and the relative importance of these duties for a given job.

Why do we do job analysis?

- Defining role profiles/job descriptions/person specifications
- Job sizing; job analysis can help determine the overall size of a role and therefore the appropriate pay grading required for it
- Developing a framework of criteria for assessment e.g. skills potential

Good Job Analysis leads to:

- Things that can be defined clearly
- Measurable concepts

Developing a framework of criteria for assessment e.g. skills potential

Less effective Job Analysis leads to:

- Loosely defined behaviors/skills which cannot be measured easily
- Behaviors/skills which cannot be measured easily

Common Methods of Job Analysis

Traditionally, job analysis was very time consuming and involved methods to collect information from multiple sources.

Structured interviews:

- Job holders can be interviewed about important behaviors required to be effective in their role, e.g. Critical Incident Technique prompts an individual to explain the positive or negative impact of an action on a specified outcome
- Line managers can also be interviewed to establish the requirements to perform well in a given role, e.g. Repertory Grid Comparisons can be used to compare skills potential in terms of their importance for a job
- Visionary interviews can be conducted in a structured way with a mixture of stakeholders to establish the key requirements for a role going forwards

Job content reviews:

Another method of job analysis is job content review. Reviewers analyze what is important for a given role by studying the job via different methods that can include

- Diaries
- Observing the job
- Doing the job
- Task/job analysis questionnaires
- Validation research

Validation research

Another method of job analysis is job content review. Reviewers analyze what is important for a given role by studying the job via different methods that can include

- Large samples of job holders or applicants
- Establishing statistical links between test scores and job performance

Methods like these, including structured interviews, focus groups and visionary interviews can now also be supplemented with much faster, online methods such as the Saville Assessment Job Profiler, a multi-rater assessment or in-person or online card sort exercises. Using these methods in combination can be much more resource friendly as they are less time-consuming.

Job Profiler and Card Sort

The Saville Assessment Job Profiler tool and the Wave Performance Culture Framework Card Sort can be used to supplement different job analysis methods.

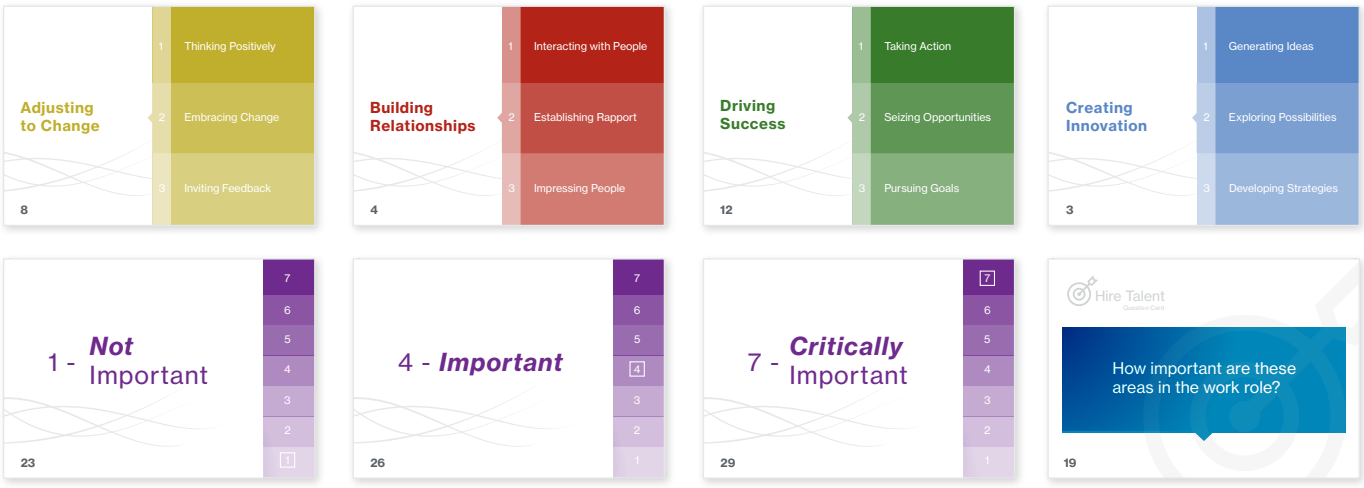
Saville Assessment Job Profiler

- Job Profiler is an online tool that takes 15 minutes to complete
- It can be used to survey different stakeholders within an organization on the importance of different behaviors and aptitude areas to a given role
- Stakeholders are asked to rate 36 behaviors and 6 aptitude areas on a 1 – 7 scale from Not Important to Critically Important, giving an overview of which areas are most relevant to the job in question. The resulted job profile aggregates the views of all stakeholders together to provide key guidance on which behaviors to assess and which aptitude areas should be evaluated using suitable aptitude assessments
- Stakeholders can also leave free-text comments on what they think is crucial to performing well in a given role

Saville Assessment Card Sort

The Hire Talent Card Deck includes: Behavior cards showing the section and dimension levels of the Wave Performance Culture framework, Ability cards showing the dimension and facet levels of the Wave Performance Culture framework, Scale cards providing structure to rank each indicator's level of important and a Question card providing direction for card sort exercises.

Using a card sort activity, stakeholders are encouraged to discuss and identify all performance indicators using 12 Behavior section cards and six Ability dimension cards. Subsequently, the Question and Scale cards can be used to facilitate further discussions of the level of important of each indicator, and to confirm the selection of relevant aptitude assessments from the Saville Assessment portfolio.



Case Study: Job Analysis

You are required to design an assessment process for the following vacancy:

- Business Development Manager

The full job description and company profile can be found on page 17 and 18. Before you design your process, you'll need to do some job analysis. Normally, you would conduct job analysis using a number of different methods involving a number of different stakeholders. Card sorts are a useful way of quickly gathering opinions from individuals or groups. Have a go at one now yourself to design your person specification. Use the steps listed to help you.

- Review your job description
- Use the Wave card deck to identify up to eight key skills potential areas (five behaviors and three abilities)
- List your key skills potential areas in the space below

Key Skills Potential Areas:

1.

2.

3.

4.

5.

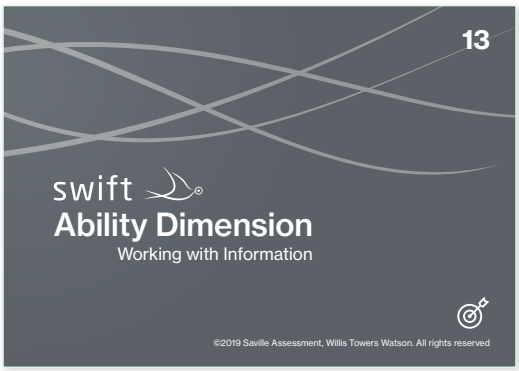
6.

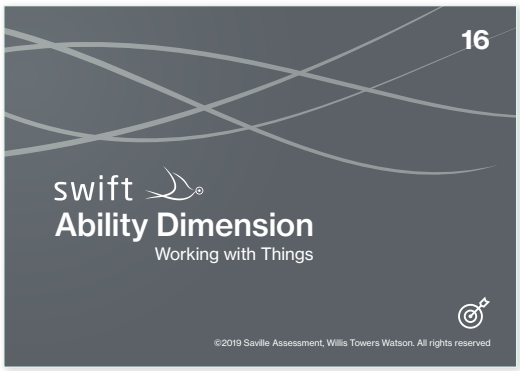
7.

8.

Hire Card Deck - Behavioral Sections

Card Deck - Ability Sections





Working with Words
Verbal Aptitude

A	Understanding Word Meaning
B	Comprehending Text
C	Making Verbal Inferences
D	Evaluating Written Materials
E	Comparing Arguments

13

Working with Systems/Logic
Diagrammatic Aptitude
Abstract Aptitude

A	Understanding Logical Rules/Sequences
B	Comprehending Process Diagrams/ Processes
C	Identifying Causes/ Rules
D	Finding Faults
E	Comparing Flowchart Sequences

16

Working with Numbers
Numerical Aptitude

A	Understanding Tables
B	Comprehending Graphs
C	Making Numerical Inferences
D	Evaluating Quantities
E	Comparing Data

14

Working with Designs
Spatial Aptitude

A	Estimating Lengths and Angles
B	Recognizing Rotated Shapes
C	Visualizing 3D Objects
D	Inspecting Objects
E	Designing Things

17

Working with Details
Error Checking Aptitude

A	Checking Letters and Text
B	Checking Numbers and Tables
C	Checking Codes and Symbols
D	Identifying Mistakes
E	Classifying Information

15

Working with Equipment
Mechanical Aptitude

A	Understanding Mechanical Problems
B	Comprehending Physical Principles
C	Estimating Movement of Objects
D	Using Tools
E	Operating Machinery

18

Job Description

Business Development Manager

A new Business Development Manager is required to head up the e-Learning Account Management Team. The role will focus on overall management of the team and supporting them in developing their existing client accounts as well as encouraging new opportunities. The Business Development Manager will inspire the team to come up with innovative e-learning approaches to provide new solutions for clients.

Key Responsibilities:

- Managing the team and coordinating their sales and account management activities
- Forming strategies on developing e-learning's usage with existing accounts and generating and following up new leads
- Generating innovative ideas and creative approaches to e-learning with due consideration of customer needs
- Providing additional training to the team to increase sales revenues
- Managing challenges encountered by the team and advising on the best course of action
- Developing and delivering effective solutions for clients
- Producing monthly billing reports for the Management Team and managing project budgets
- Analyzing and reporting on solution effectiveness

Required Skills and Experience:

- Proven sales track record
- Influencing and negotiation skills
- Interpersonal and communication skills
- Able to network and build relationships with a range of individuals
- Excellent project management skills
- Able to motivate a team to achieve targets
- Able to develop innovative approaches to meet business objectives
- Can adapt to challenging situations and remain positive
- Approachable, providing support and sharing expertise with the team
- Previous experience working with dynamic simulation software and knowledge of e-learning programs
- Strong written & verbal communication skills
- Strong numerical & logical thinking skills

Company Overview

Company Profile: Specialists in developing new digital media technology. Experts in developing virtual simulations, marketing and advertising campaigns, online training programs and applications for mobile devices. Due to the strong growth in the e-learning industry and solutions which have proved to be very popular with clients, Tradigital is fast becoming a market leader within the e-simulation and application industry.

Number of Employees: Approximately 400.

Vision: Delivering high quality simulation solutions which educate, inspire and captivate our customers.

Latest News: In order to meet the demand and develop opportunities with new and existing clients, Tradigital have created a new Account Management Team. The team is tasked with increasing revenues from existing clients, and identifying and converting new sales opportunities.

The Account Management Team aims to:

- Identify and successfully secure sales with new clients
- Manage a portfolio of key clients, supporting the implementation of e-learning sales projects
- Provide ongoing support to develop business opportunities within these clients

Account Managers need to liaise closely with the Marketing Team to initiate and manage promotional campaigns and with the IT Development Team who develop the software to the client's specifications.

The Account Management team consists of 14 individuals who were previously Sales Advisors at Tradigital.

Current Situation: There is a need to appoint a Business Development Manager to head up the newly created Account Management Team.

Notes

Profiling Requirements

Ensuring that the correct assessments are used in a particular context is extremely important. Choosing the correct assessment helps to increase the reliability, validity and fairness of any assessment process, and also guards against risks associated with incorrect or poor assessment use. One of the most important stages in this process is the profiling of the requirements for an assessment process, i.e. knowing what you're trying to measure and choosing the best tools which help you to do this.

Wherever possible, questionnaires and tests should be chosen on the basis of a thorough job analysis to ensure that decisions are being made with the use of relevant information. Job analysis is a process to identify and determine in detail the particular duties and requirements in a role, as well as the relative importance of these for the job. This can include deciding which aspects or scales from an assessment will be considered relevant to the job. When looking at the importance of Wave scales for a particular role, selecting six Wave sections as critical areas is generally a realistic and manageable number.

Job analysis is an essential pre-requisite to choosing which psychometric tests and questionnaires to use. There are several ways to profile a role, including: interviews with incumbents and supervisors, questionnaires (structured, open-ended, or both), observation, and gathering background information such as job descriptions. It is common to use more than one of these methods.

Saville Assessment have developed the Job Profiler and Wave Performance Culture Framework card deck to support organizations in their job profiling activities.

Wave Performance Culture Framework

The Wave Performance Culture Framework card deck can be used for a wide variety of applications. The cards cover Behavior, Ability and Global measures from the overarching Wave Performance Culture Framework.

The framework gives enormous flexibility to measure performance and work culture. It provides a 'language of work' that helps workers and managers describe work, performance, and culture in a clear, concise and objective manner.

You can decide how to classify and prioritize work elements, from a very broad level through to a very detailed and granular level. That flexibility allows users to focus in at the most appropriate level for their application and to 'drill down' when more specific information is needed.

The card decks can be used on a one-to-one basis, with small groups/teams or with large focus groups as part of interactive sessions to assess key characteristics. It offers an engaging and interactive approach with line managers and non-HR teams at all levels in an organization. The vocabulary is simple, direct and jargon-free. The cards enable users to cover a lot of ground quickly and tease out areas of agreement/disagreement using a constructive and non-threatening process.

Job Profiler

Saville Assessment have developed the Job Profiler questionnaire, an online measure (taking just 15 minutes to complete) that captures the essential features of jobs in an efficient and effective manner.

The Job Profiler includes multi-ratings, gaining perspectives from the job holder, boss, stakeholders and reports.

The assessment covers Behaviors, Ability and Global measures from the overarching Wave Performance Culture Framework.

The results provided by each rater group are represented by a different shape and positioned on the rating scale with arrows reflecting any differences or ranges in opinion, as shown on the next page.

There is also a free text section that adds richness to the data gained from the rating scale. The free text allows users to explore opinions of key skills and knowledge from different rating groups in more detail to help identify the core role requirements.

Please speak to your course director if you'd like to discuss job analysis and job profiling requirements in greater depth.

Module 3: Assessment Choice and Administration

Considerations for Choosing Assessments

Early Considerations

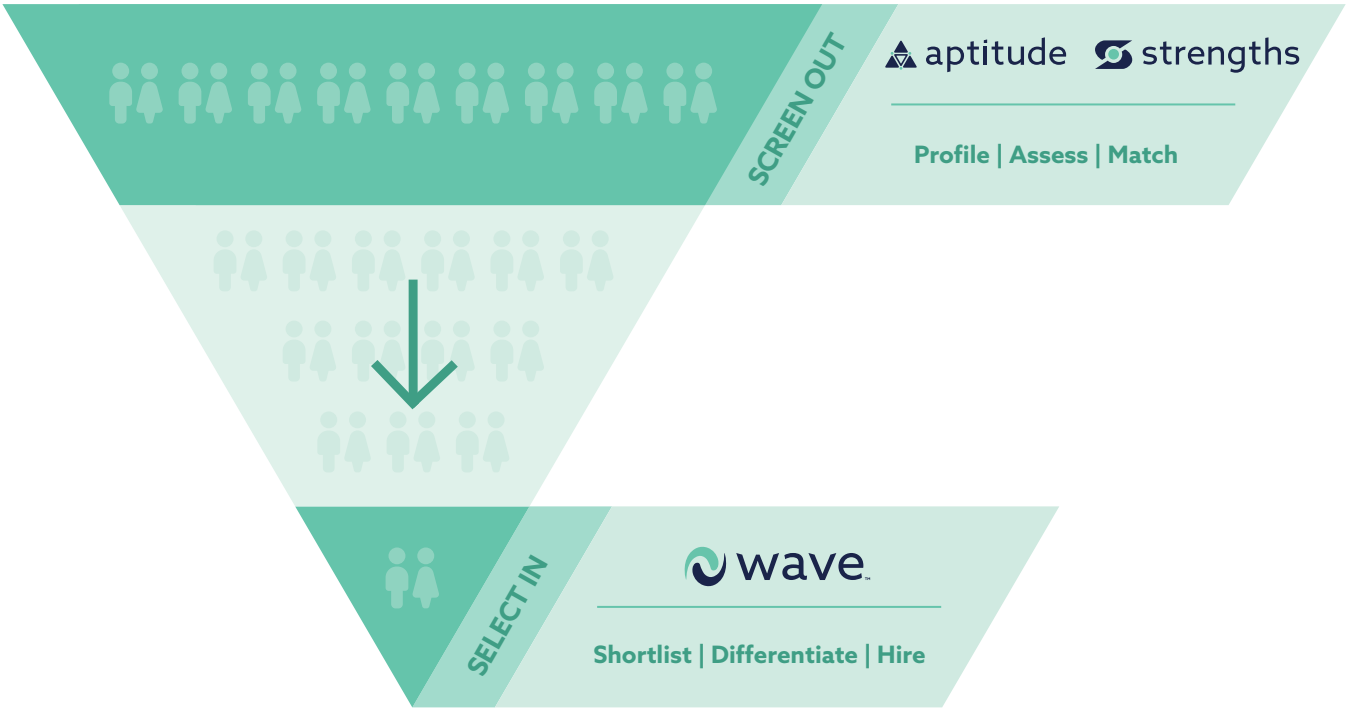
- Do the test yourself
  - Does it look good?
  - Does it make sense?
  - Is the content relevant to the role?
  - Does the content appear fair and inoffensive?

Other considerations (to be discussed later)

- Is it psychometrically sound?
- How much does it cost in total?
- What are the administration practicalities (screening online, number of candidates, etc.)?

Notes:

Screen Out, Select In



Notes:

Our Methods of Screening

- Aptitude tests
- Behavioral screening questionnaires
- Language tests, e.g. Workplace English
- Situational Judgment Tests

Saville Assessment Aptitude Tests

Analysis Range			
Target Group	Aptitudes Assessed	Norms Available	Test Options
Directors Managers Professionals Graduates Management Trainees	Verbal (V) Numerical (N) Diagrammatic (D) Abstract (A) Recall (R)	Senior Managers & Executives Professionals & Managers Graduates Mixed Occupational Group	Online Unsupervised Combined Tests - Swift Executive Aptitude (V, N, A) - 18 mins - Swift Global (R, N, A) - 12 mins - Swift Analysis Aptitude (V, N, D) - 18 mins - Swift Analysis Verbal & Numerical (V, N) - 24 mins Single Tests - 24 mins

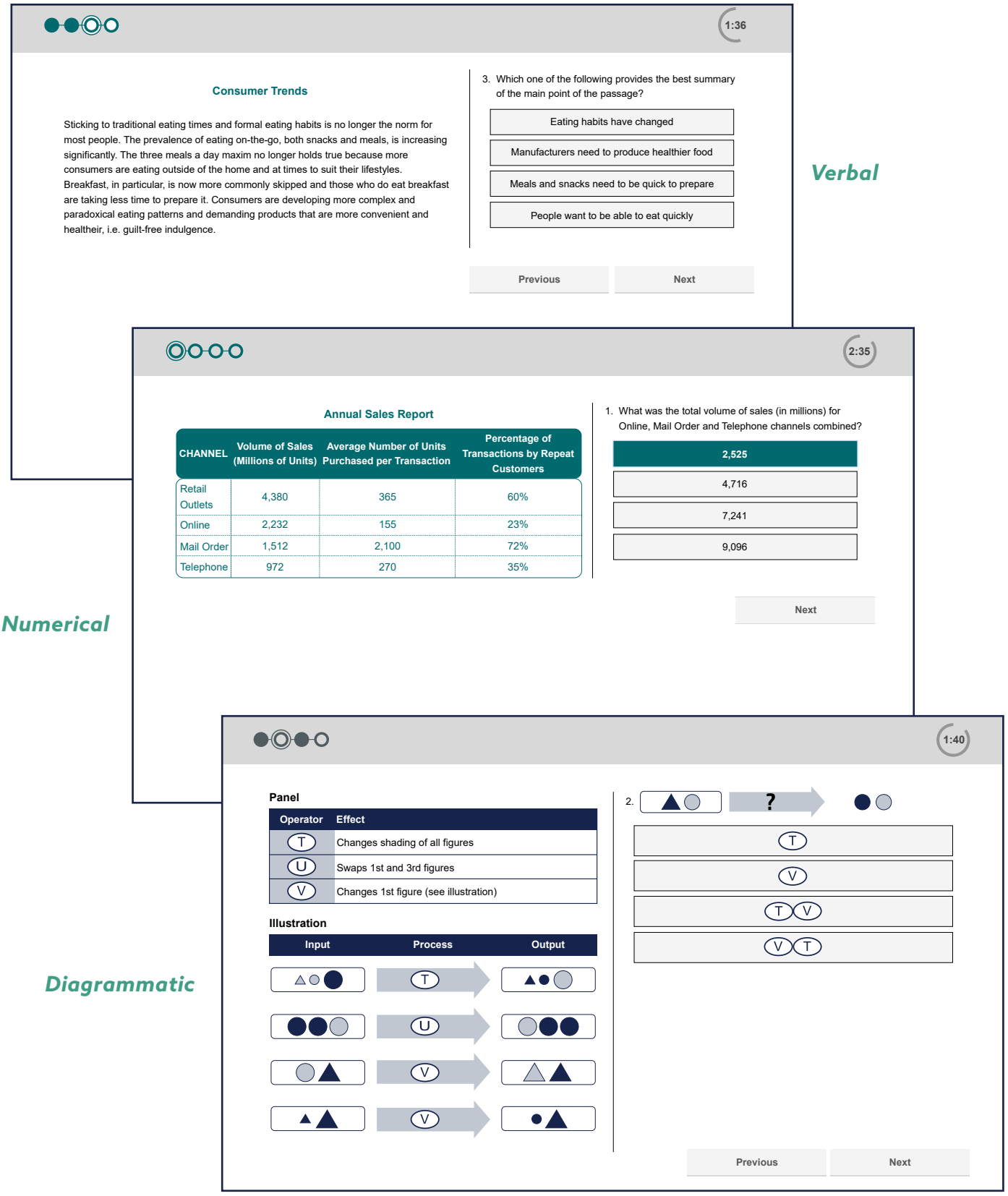
Comprehension Range			
Target Group	Aptitudes Assessed	Norms Available	Test Options
Administrative Roles Customer Service Roles Operational Roles Commercial Roles	Verbal (V) Numerical (N) Error Checking (EC)	Mixed Occupational Group Individual Contributors Apprentices Foundation (International Only)	Online Unsupervised Combined Tests - Swift Comprehension Aptitude (V, N, EC) - 9.5 mins - Swift Comprehension Verbal & Numerical (V, N) - 12 mins Single Tests - (V) 16 mins, (N) 16 mins, (EC) 6 mins

Technical Range			
Target Group	Aptitudes Assessed	Norms Available	Test Options
Production Roles Construction Roles Engineering Roles Scientific Roles	Spatial (S) Mechanical (M) Diagrammatic (D)	Mixed Occupational Group Apprentices	Online Unsupervised Swift Combination - 10 mins Single - (S) 8 mins, (M) 12 mins, (D) 16 mins

Swift Apprentice Aptitude			
Target Group	Aptitudes Assessed	Norms Available	Test Options
Apprentices	Verbal (V) Numerical (N) Error Checking (EC) Spatial (S) Mechanical (M) Diagrammatic (D)	Apprentices	Online Unsupervised Swift Combination - 19.5 mins

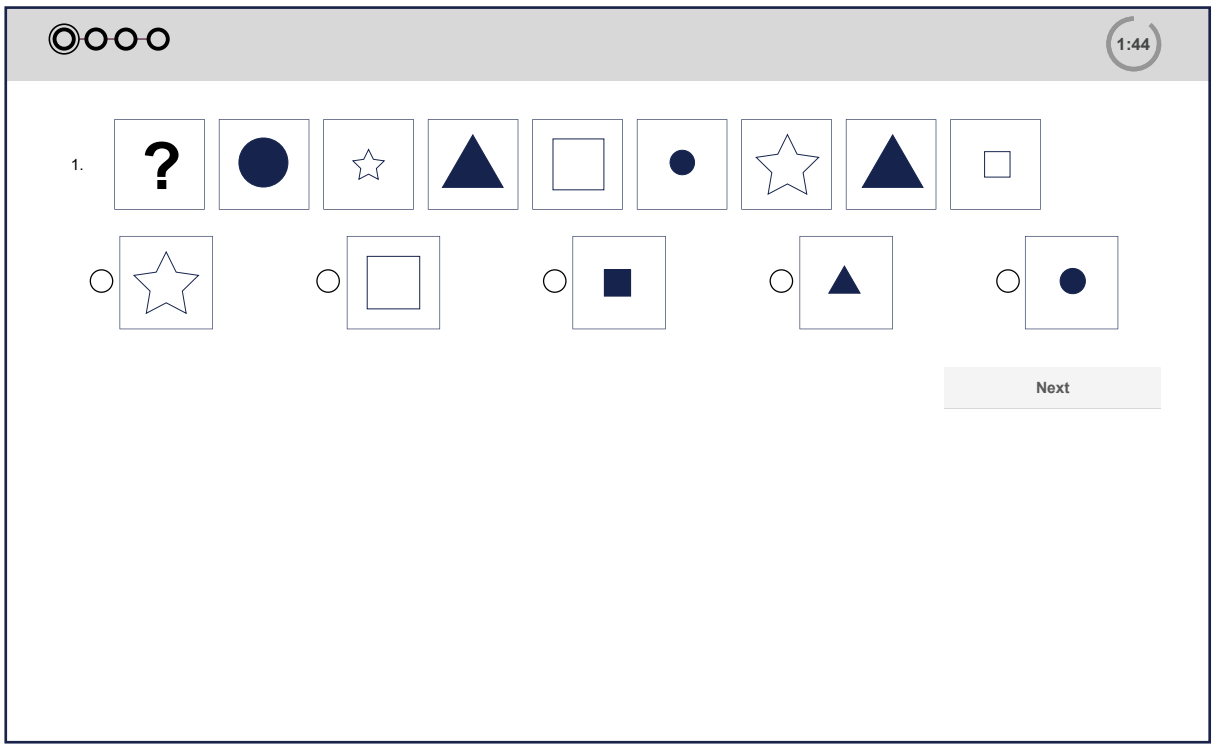
Analysis Aptitudes

Swift or single aptitudes



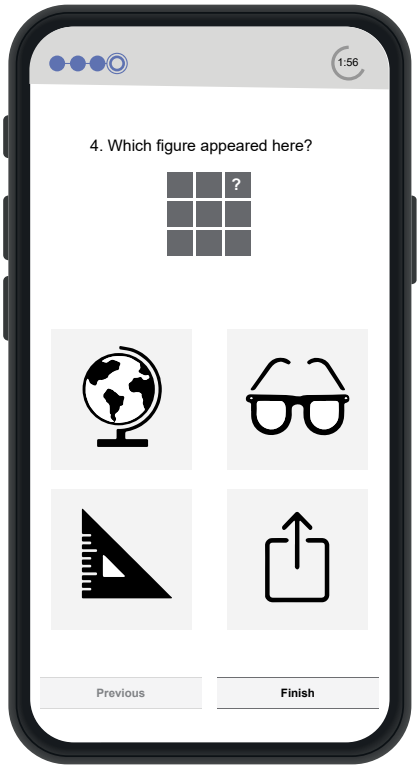
Abstract Reasoning Aptitude

Swift Executive or single aptitude



Swift Global

Swift



Comprehension Aptitudes

Swift or single aptitudes

1:34

Office Accident Book

Basic Advice on First Aid at Work

The following information should be recorded for any incidents involving injury or illness that have occurred inside the office building:

- Date, time and place of incident.
- Name and job of the injured or ill person.
- Full details of the injury or illness and any first aid given.
- What happened to the casualty immediately afterwards? For example, did the casualty return to work, go home or to hospital?
- Name and signature of the person dealing with the incident.

This information will help identify accident trends and possible areas for improvement in the control of health and safety.

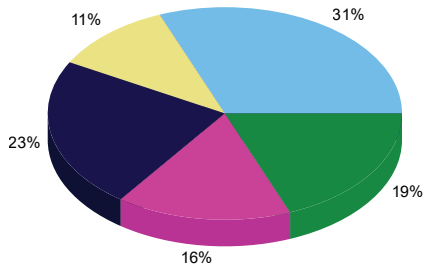
1. Which **one** of the following would best replace identify while maintaining the meaning of the passage?

- name
- designate
- disassociate
- spot

Verbal

1:46

Monthly Review of Staff Time by Activity
(Total Number of Days: 2000)



1. What percentage of time was spent on Training and Administration?

- 16%
- 19%
- 25%
- 35%
- 42%

Numerical

0:34

Subscription Details

Company Name	Serial Number	Type*	Date
Account Manager Services	114321	G	03/05
Inavat Solutions	719909	N	09/09
Office Experts	132461	N	12/10
Advanced Solutions	910073	P	08/11
Softwarehouse	100333	G	01/10
Marketing Focus	618736	P	06/08

May June July

***Coding Key for Type:**

P = Personal **N** = National **G** = Global

Web-Log **SoftCo**

Company Name	Serial Number	Type*	Date
Advanced Solutions	910073	Personal	08/11

2. Please select the option, or options, which apply.

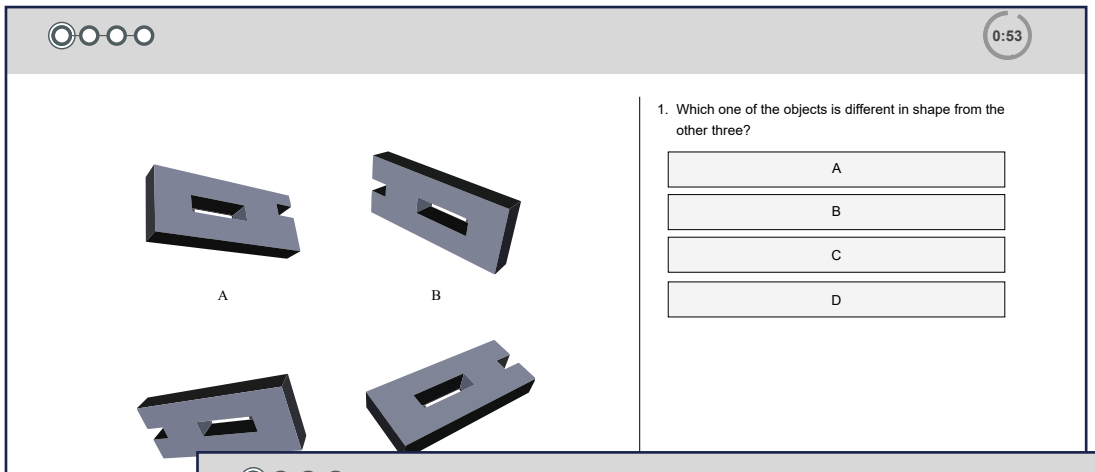
- The entire item is correct
- There is an error in the Company Name
- There is an error in the Serial Number
- There is an error in the Type
- There is an error in the Date

Previous Next

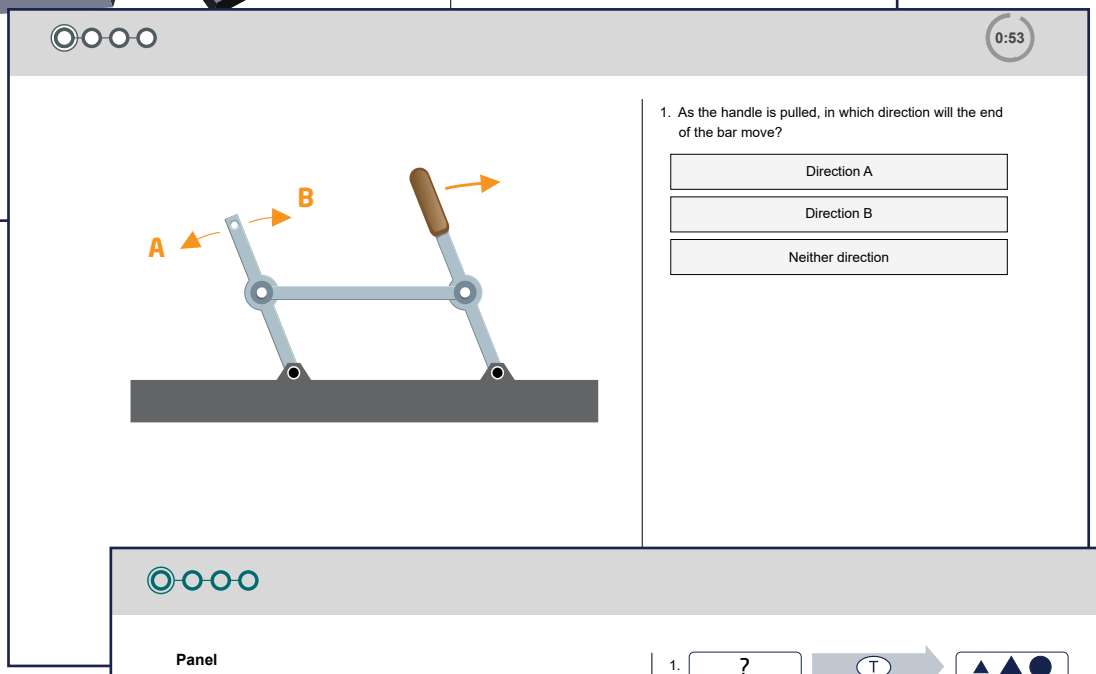
Error Checking

Technical Aptitude

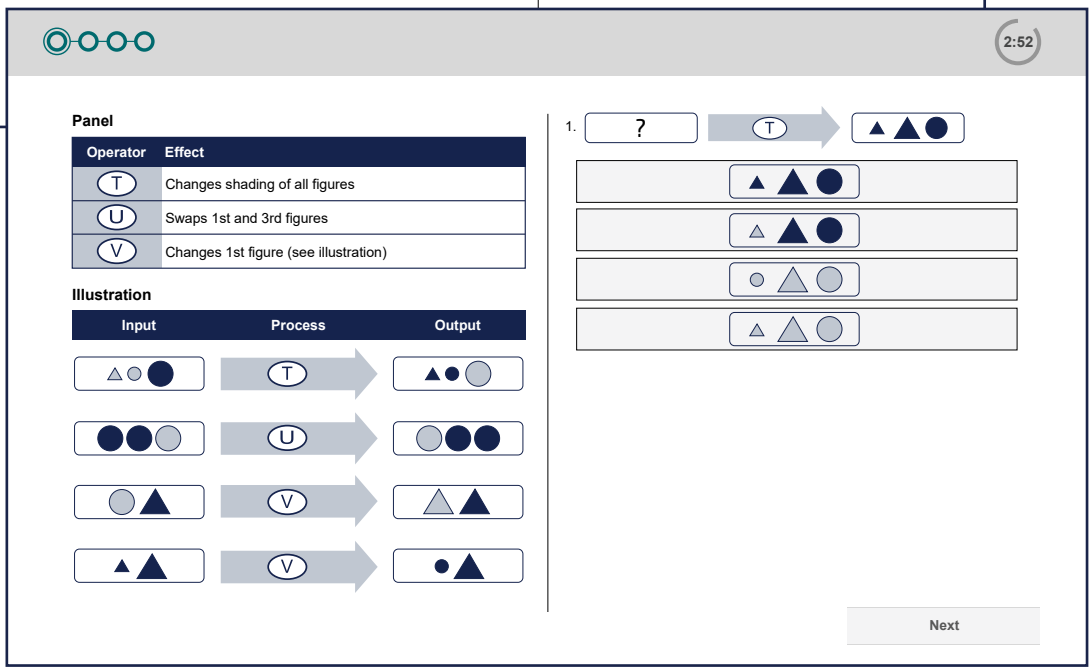
Swift or single aptitudes



Spatial

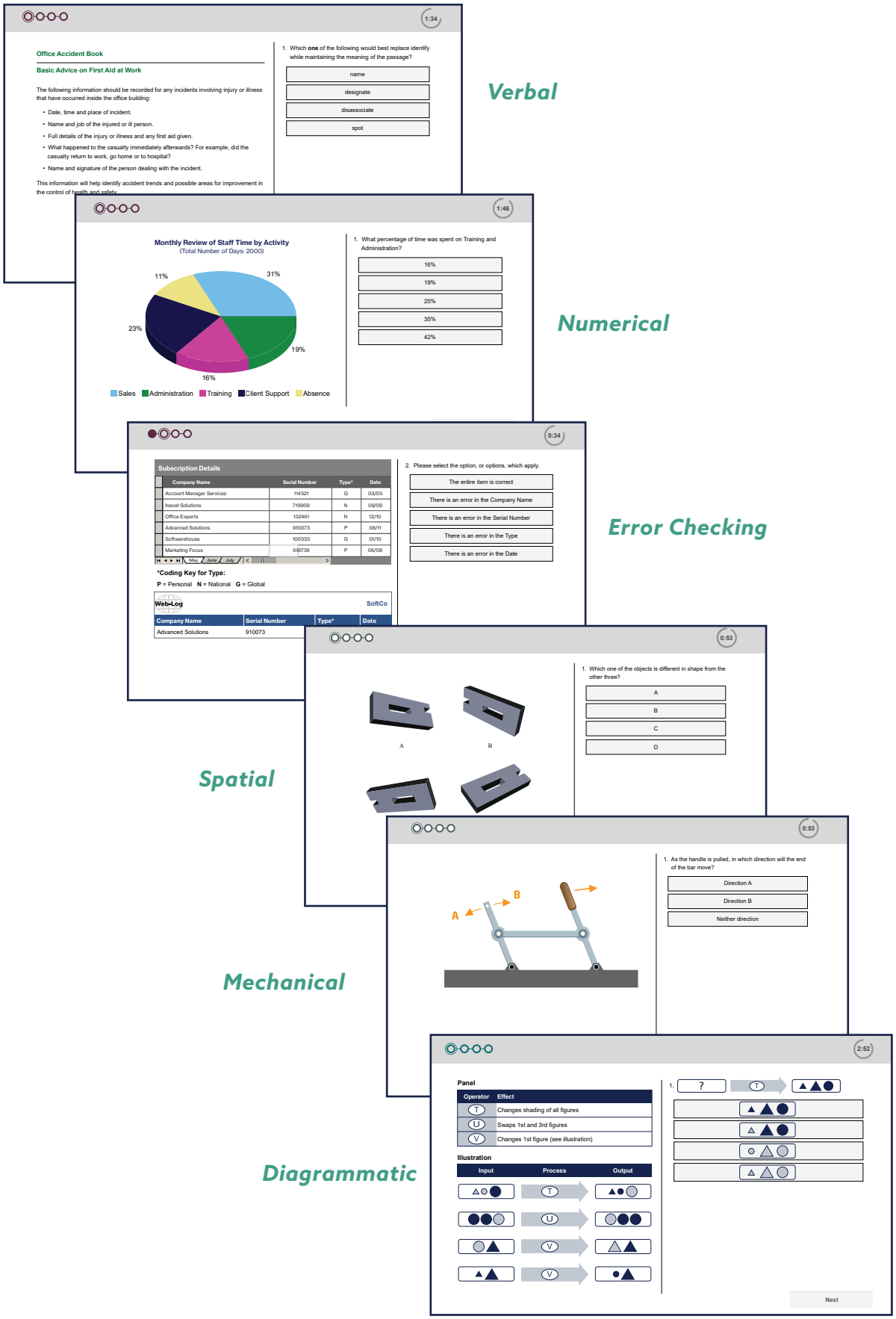


Mechanical



Diagrammatic

Swift Apprentice Aptitude



Why Use Cognitive Ability Tests?

Hire

Cognitive ability tests are mostly used for recruitment purposes, either in screening or selection

Build

Test are used less frequently for individual development, although career guidance and planning tools often contain an ability component

They do predict training performance!

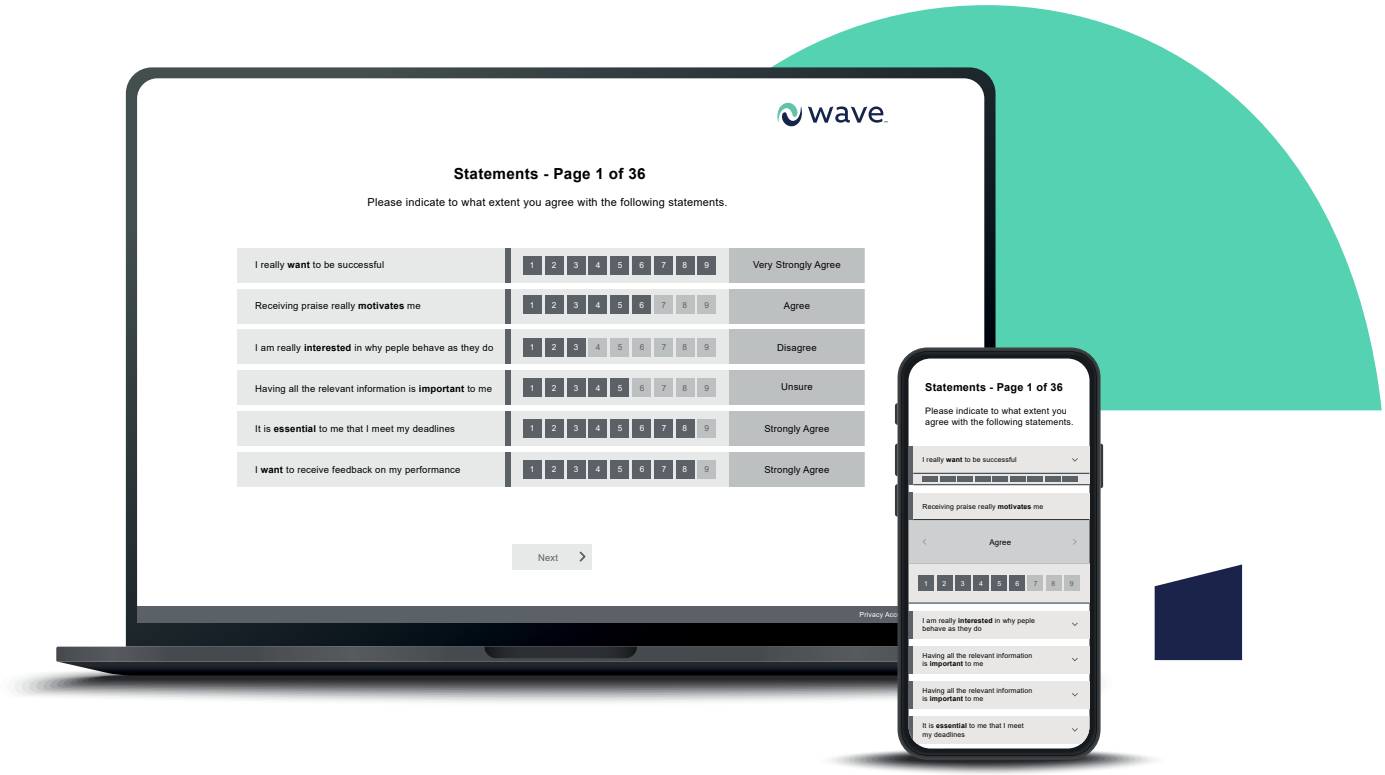
Lead

Despite many leaders' avoidance of testing, cognitive ability has been shown to be especially predictive of performance at senior levels

Notes:

Behavioral Questionnaires

- Underpinned by the highly-researched Wave model
- In-depth Styles assessments to explore working styles including exploring situational differences and discrepancies between motivation and talent
- Short behavioral screening assessments that can provide one fit score for rapid decision making in screening
- Mobile-first responsive design for an improved candidate experience



Workplace English

- Assesses candidates' ability to understand workplace English
  - Secure and easily administered online in less than 20 minutes
- Tests for five different industry sectors:
  - Customer Service, Healthcare, Hospitality, Office, Operations
- Each test comprises 33 items:
  - 24 industry-specific items
  - Nine general workplace items

Example question

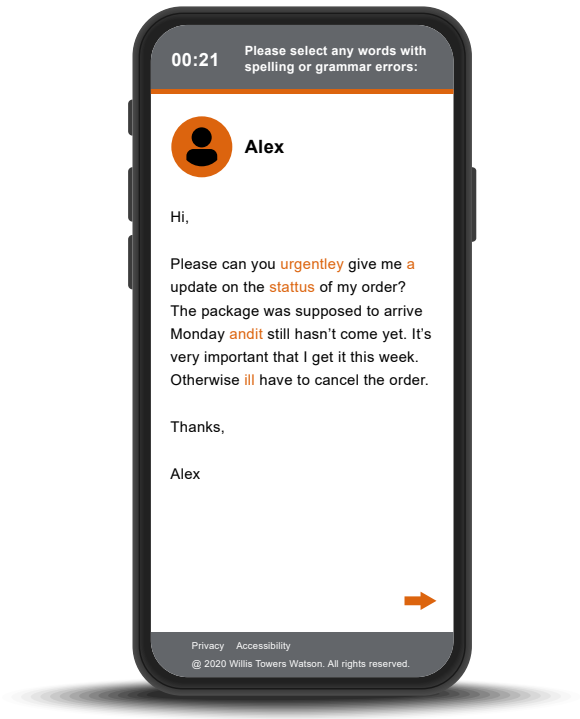
I [?] review the situation tomorrow

1. am

2. will

3. do

4. think



Check

- Customer Check is an online assessment identifying how well individuals will interact with customers and represent your brand on live web chats
- Highly-relevant assessment specifically designed to replicate typical customer contact
- Compatible with a range of devices, including laptops, tablets and mobiles
- Quick and engaging assessment with a 10-minute completion time

Notes:

Situational Judgment Tests

- Candidates are presented with workplace scenarios and different responses to the scenarios
- Candidate are required to respond on a 1 - 7 scale to rate how effective each response is
- This evaluates how suitable candidates' judgment is in work-relevant tasks
- Candidates receive a feedback report outlining the effectiveness of their judgment in the given areas

Notes:

Assessment Choice

Having completed your card sort you can now start to design your assessment process.

Complete the table below, listing your eight key skills potential areas and identify how you might measure the required characteristics.

Job Title - Business Development Manager

Person Specification (Key Skills Potential Areas from Card Sort)	**Sift Out** (for example; Situational Judgment Test, Behavioral Assessment, Ability Tests)	**Select In** (for example; Structured Interview, in-Person Assessment Exercises)

Example Selection Process

Combined Assessment Approach

When planning a selection process consider:

- The number of applications you expect to receive
- The number of positions you need to fill
- The level of the role
- What is being asked of the candidate at each stage
- Logistics and practical considerations
- Screening out and selecting in
- Whether there are any reasons why a test should not be used or should be restricted*

Going forwards you can use Saville Assessment's brochures, handbooks and the notes from this workbook to support you choosing psychometrics and planning your process.

*E.g. not appropriate for the age group, educational level, reading level, or even that the test contains content that is known to one particular cultural/ethnic group and not others.

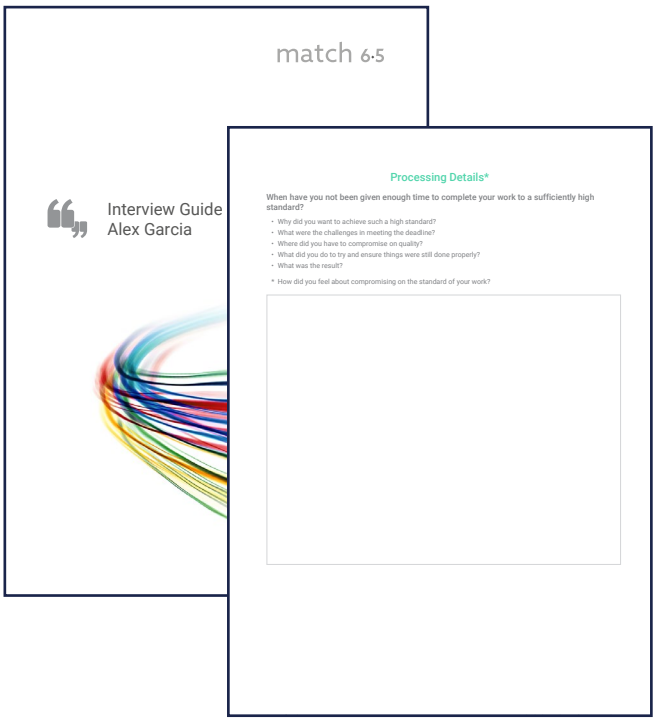
Using a Fit Score Approach

- Focus on creating one engaging event rather than managing a number of stages.
- Collect all the data you need to help you make an accurate, objective and fair decision.
- A simple change can help you achieve against all of the areas of return on investment; quality, efficiency, cost, DE&I, engagement.

Role Fit available from Match 6.5 and Aptitude combined

First Name	Last Name	Combined Fit Score	Behavioural Fit Score	Aptitude Score	Norm Group
Lee	Crouch	38.7925	41.79	38	Individual Contributors (INT, IA 2022)
Constance	Markievica	40.0393	37.6883	54	Individual Contributors (INT, IA 2022)
Sabrina	Smith	44.2315	46.7398	41	Individual Contributors (INT, IA 2022)
Jooris	Axelstein	45.1217	46.3501	45	Individual Contributors (INT, IA 2022)
Isabel	Mebarak	48.0183	47.1387	52	Individual Contributors (INT, IA 2022)
Lucia	Fernando	48.6593	46.1332	57	Individual Contributors (INT, IA 2022)
Chris	Park	54.5712	53.9965	53	Individual Contributors (INT, IA 2022)
Sam	Jenkins	59.5189	57.0387	60	Individual Contributors (INT, IA 2022)
Faisal	Wooton	62.219	62.0572	54	Individual Contributors (INT, IA 2022)

Interview Guides and Candidate Reports Powered by Screening Assessments



Assessment Choice

Evidence supports only certain exercises measure classic behavioral skills potentialss

- Interview questions focusing on skills potential
- Behavioral/personality assessment
- Aptitudes load onto certain criteria such as ‘Solving Problems’

Aptitude tests have good validity so are a good predictor of performance

Mechanical validity where clear decision rules and cut-offs are applied consistently outperforms human judgment in screening every time!

- Generally we would advise that hiring managers should be given pre-qualified applicants without access to previous screening assessment scores

Assessment center exercises should be used sparingly as have low validity

- Exercises should lead to one exercise score, not separate skills potential scoress

Situational Judgment Tests are engaging, reflect an organization’s brand and offer strong validity, despite not measuring skills potential (they measure judgment)

Administration Formats

- **Unproctored** – This 'controlled mode' provides secure access to specific individuals to complete online tests in an unproctored setting e.g. a password and link are emailed

  Security advice for invited access administration:

  – Ensure candidates are pre-qualified and cannot sit a test multiple times

  – Use the advice on test resets documented in your organization's testing policy

  – Use item-banked tests, e.g. gradient step testing

  – Review candidate response patterns ('forensic analysis'), e.g. unusually fast response times

  – Retest under proctored conditions

- **Proctored** / **Remote Supervision** – This 'managed mode' is where a trained administrator verifies and oversees the test environment/process throughout, e.g. at a secure testing center

Notes:

Overview of Assessment Administration

Key Considerations:

- Select appropriate tests based on job analysis
- Review your organization's testing policy to ensure that your communication and administration are compliant with it.
- Prepare your communication (email/letter):
  - Explain which test(s) they are being asked to complete and the timing of each
  - Explain why the test is used as part of the process
  - Check for any reasonable adjustment requirements and any anticipated problems completing the tests
  - Ensure candidates have access to preparation/ practice materials
  - Inform candidates of how you will store their data and for how long, and who will have access to them (i.e. adhering to applicable legislation)
  - Inform candidate of next steps, e.g. when they will receive feedback secure testing centers

Notes

Saville Assessment Cognitive Ability Test Types

There are two types of Saville Assessment Ability Assessments:

Single Ability Tests: In-depth single tests assessing one ability only. Single tests consist of around 28 questions and test timings range from 6 to 24 minutes, depending on the ability measured. Error Checking and Spatial reasoning tend to have shorter test times, with Verbal, Numerical and Diagrammatic tests having longer test times.

As well as overall scores, users can profile scores achieved on particular types of items in the test; these are presented as 'sub-scores' on test profile charts, giving richer and more detailed information on performance.

Single tests are developed for use in both unproctored and proctored, online situations. For unproctored administration (Invited Access – IA), single tests are available for each of the three ranges: Analysis, Comprehension and Technical. These tests are globally applicable across industries. Single tests for use under proctored conditions are broken down further, by key target work groups:

Swift Ability Tests: Combination tests assessing multiple abilities using significantly shorter testing times than with separate single ability tests. A Swift test includes around 24 items with test times ranging from 9.5 to 24 minutes. Swift tests provide a mechanism to quickly establish a candidate's overall ability across different measures. These tests are available for use in both unproctored and proctored, online situations.

Swift Ability tests are aligned to the three ranges: Analysis, Comprehension and Technical. They are globally applicable across industries and do not break down further by work group.

Saville Assessment Ability Test Ranges

There are three main ability ranges in the Saville Assessment test portfolio: **Analysis, Comprehension and Technical**. The Analysis range has been designed for use with directors, managers, professionals, graduates and management trainees. The Comprehension range has been designed for use with operational roles in manufacturing, engineering, construction and transport, lower level commercial roles in sales, marketing, business development and financial services, customer roles in call centers, hospitality, leisure, health and education and administrative roles in private and public sector offices. The Technical range has been designed for use with production, construction, engineering and scientific roles.

Saville Assessment Work Strengths

Saville Assessment Work Strengths is a short but highly valid assessment designed for use in volume recruitment across a wide range of different job roles. Work Strengths provides a highly positive approach to the selection of staff, with feedback focusing on candidates' strengths.

Work Strengths: suitable for use with graduates, management trainees, managers and professionals

The tool is based upon research integrating personality, skills potential and overall effectiveness at work and is aligned to both the Big Five personality facet model and the Great Eight Performance model. Its development benefited from a performance-driven approach, whereby the validity of the questionnaire is maximized by selecting items from the pool that are most predictive of performance at work.

The assessment uses a dynamic response format that utilizes both ratings and rankings, allowing for identification and control of distortion.

The Work Strengths output is simple to use, enabling managers across organizational levels to interpret profiles easily and accurately. It provides feedback on work culture and the environments in which a candidate is likely to be most and least suited, as well as optional interview questions.

Match 6.5

Match 6.5 is a fast and valid tool for screening high volumes of candidates; taking just 6.5 minutes to complete. It allows the client to screen with one behavioral fit score alone or a fit score combining behavior and aptitude.

Match 6.5 easily screens across different roles for large organizations hiring across different job areas.

Saville Assessment Workplace English Tests

Workplace English tests assess an individual's ability to understand workplace-relevant sentences in English. All tests are available online for unproctored administration. Separate online versions are also available for proctored testing.

Saville Assessment Situational Judgment Tests

Situational Judgment Tests or SJTs provide engaging, realistic, work-related previews of the role by presenting candidates with scenarios they are likely to come across on the job. Candidates are then presented with a series of response options for each scenario and asked to rate the effectiveness of each. SJTs create opportunities for impactful employer branding and offer fast and effective screening for high volumes of candidates.

Customer Check

Customer Check assesses candidates' ability to identify errors in customer service message exchanges. Designed for candidates in customer service roles to assess how well individuals will interest with customers and represent organizational brand on live web chats.

Notes:

Administration Guidelines

Preparing the Session:

- Select appropriate tests
- Timetable of the day
- Inform the candidate: invitation letter/email, check accommodations and reasonable adjustments, preparation guides
- Materials – order in, check re-usables, if you are using hard-copy materials
- Question booklets, answer sheets, pencils, rough paper, calculators, administration instructions, stopwatch, test log. If you are doing online assessments, ensure laptops, mouses, keyboards are all suitable, check internet connection if this is required
- Familiarity with test – materials, admin instructions, example questions. If you are administrating an online assessment, the example questions will be asked and explained on screen for the candidates but it is still a good idea to understand what the candidates are completing
- Check testing environment

Introducing the Session:

- Introduce yourself
- Rationale for testing
- Description of tests used
- Length of test session
- Feedback arrangements
- Confidentiality, including data storage (under applicable legislation) and who has access, i.e. authorized test users
- Informed consent
- Housekeeping points: reading glasses, phones, toilets
- Testing conditions
- Any questions

Administering the Session:

Follow administration instructions precisely!

- Distribute candidate materials but not test booklets (if using hard-copy tests). If you are using online assessments, ensure these are set up ready for the session
- Read clearly and slowly
- Ensure personal details are filled in for hard-copy. If testing online, this will be pre-entered
- Help individual candidates to get examples correct. Online, examples will be explained on screen to candidates but you can also answer questions relating to the example questions
- Use a stopwatch and time independently. Online testing will accurately ensure testing but it is a good idea to have a secondary measure in case the testing session is disrupted
- Keep your eyes on the room throughout test
- Ensure everyone has stopped. This is less important in online testing as timers will end candidate sessions at a set point
- Collect all materials; even with online testing you may have rough paper/notes to collect in
- Thank the candidates

Sample Introduction to Aptitude Assessment Session

Hello, welcome to……....., I'm……....., I am a..... with…….. I will be conducting this testing session with you.

You should have all received the details describing what will happen during the testing session. I'll be asking you to complete……..…….test(s). You will then be having an interview/completing an in-tray exercise etc. We hope to complete the assessments by…….....…….o'clock today.

Before you start the test(s), I'll explain why we're using them; we use them because they give us a fair and objective assessment of your skills in……..... .....……..(Insert short description of relevant ability). These are important skills for the role you have applied for and we find that those applicants who do well in the tests subsequently do well in the role. It is also in your own interests as well as ours that you are suited to the role for which you've applied. We also get additional information from the tests that we can't readily get from other aspects of our selection process.

Decisions on whether to progress your application are based on all the information we gather from you today; you can contact me for feedback on your test performance, and I'll give you my contact details later.

Your results from this session are confidential and will be stored in line with applicable legislation. Your results will only be shared with those involved in the recruitment process; do I have your informed consent to continue?

The whole test session will last approximately…….....……..minutes (add on 15 minutes to each test duration). The first test will last for exactly……..…minutes. Please don't leave the room once we're underway with the tests, so now is a good time to visit the toilet or collect reading glasses if you need them; please do not talk to other candidates once we've started the tests and please can you ensure mobile phones are switched off.

I'd like to take this opportunity to advise you to work quickly through the tests and try to answer as many questions as possible. I'll be reading the instructions for the test to you from a card, so please listen carefully. There are some example questions at the beginning of each test so you know what to do on the test itself. These are not timed or scored and I'll walk round to ensure that everyone has got them correct before we continue – they don't contribute to your overall score.

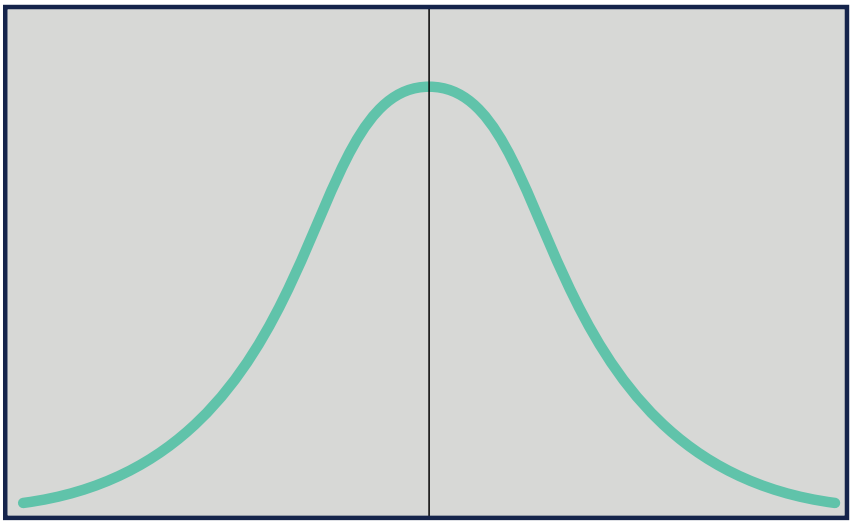
Module 4: Test Scores

Scoring

An applicant scores 19 correct (their raw score) on a numerical test (28 items)

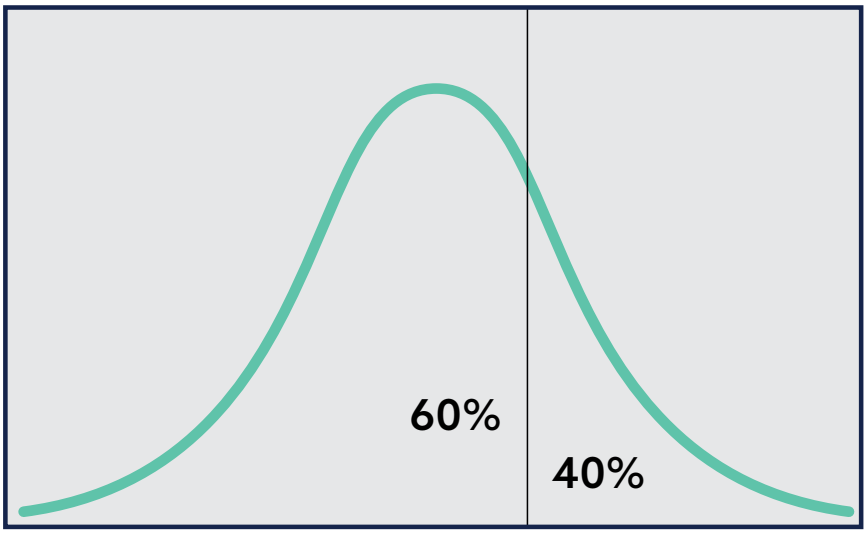
How well has the applicant performed?

The Normal Distribution



Notes:

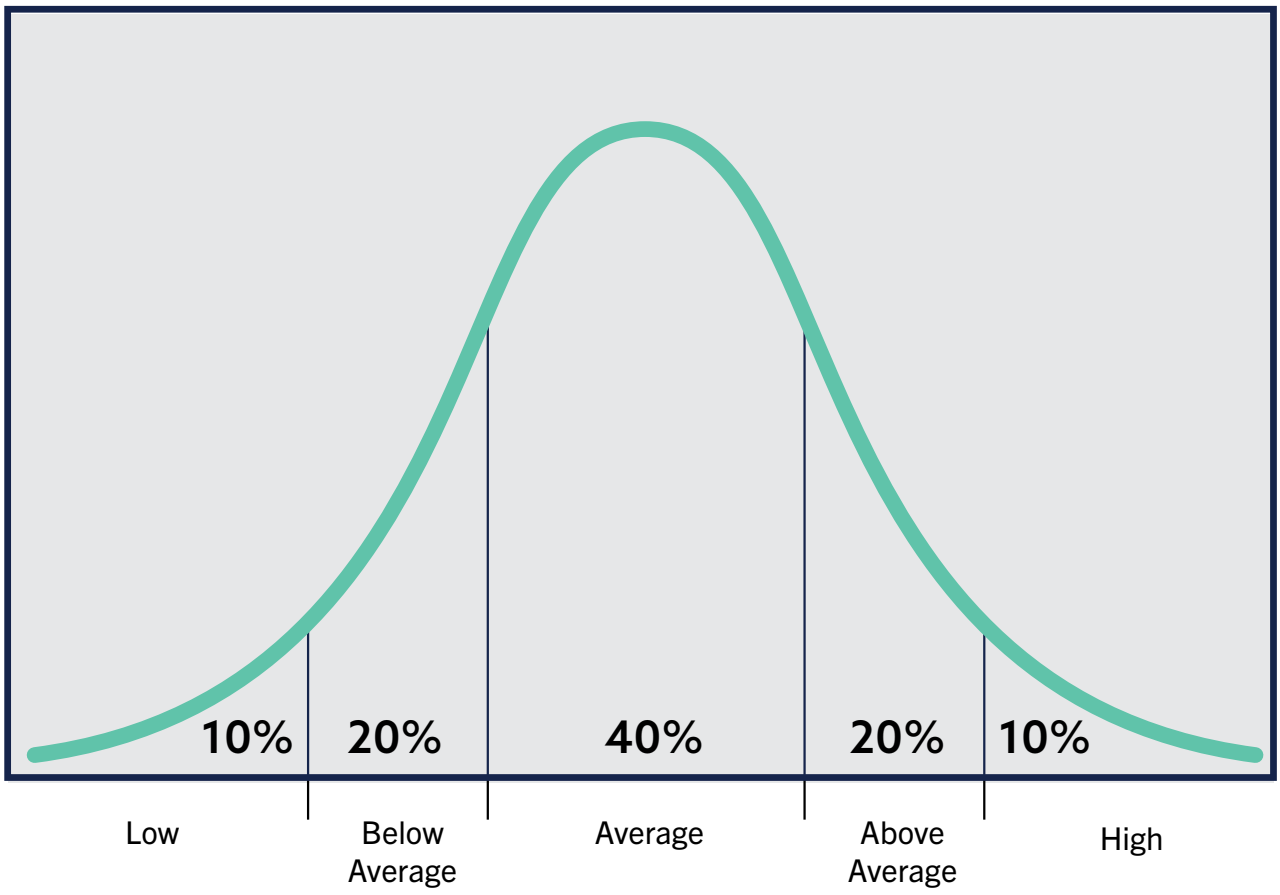
Percentiles



Describing Percentiles

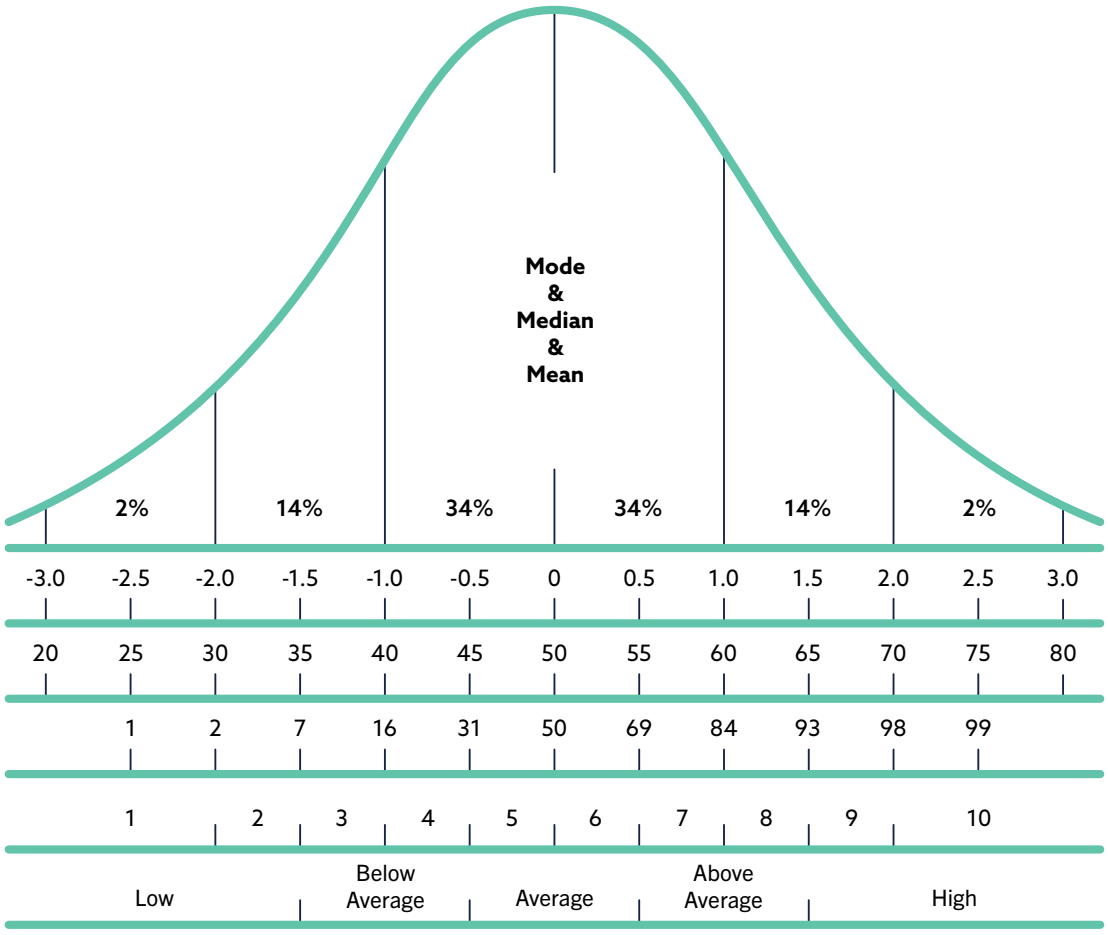
- Candidate A: 56th %ile Average
- Candidate B: 33rd %ile Average
- Candidate C: 68th %ile Average

The Normal Distribution Performance Bandings



Notes:

Ways of Describing Scores



Notes:

Selecting Test Norms

- Specific vs general benchmark norms
  - Use of specific norms which represent only one protected group, e.g. female only or white only, carries a clear risk of illegal discrimination
- Should take account of:
  - job applied for
  - educational level
  - work experience
  - generalisability / representativeness
- Suitable size - usually 150+ people
- Statistical representation of the population they are drawn from
- Individuals who represent real testing situations

Saville Assessment Aptitude Norms

Geographies

- International
- Regional (e.g. Europe, Africa, Latin America)
- Country (e.g. UK, China, Spain)

Levels/Roles

- Senior Managers and Executives
- Professionals and Managers
- Individual Contributors
- Graduates
- Mixed Occupational Group
  - Mixed Commercial, Customer, Administrative, Operational, Technical
- Apprentices
- Technical Occupations
- Foundation Level
- English as an additional language

Notes

In this section of the course notes, we will introduce you to test scores and answer the question “Why use test norms?” We will be looking at tests norms in more detail later on in the course. For now, we will become familiar with the normal distribution, a key concept in test norms and test score interpretation and understanding.

Raw Scores Vs Normed Scores

For a score to be meaningful to us, we need some way of comparing the score achieved by an individual on a test against the scores achieved by a representative/ relevant sample of people, i.e. a benchmark or norm group. To start with, consider the following example:

An applicant scores 19 correct answers (their raw score) on a Saville Assessment Professional Numerical Analysis test, which has 28 questions (also known as items) and lasts for 20 minutes. How well have they performed?

The answer is that without knowing how well the rest of the group has performed, it is impossible to know whether 19 is a good score or a bad score.

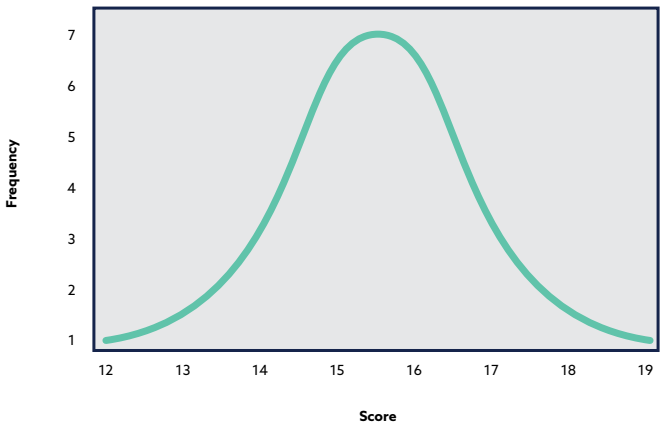
Frequency Distributions

One way to look at a group’s scores is to produce a frequency distribution. On the horizontal (y) axis, the scores on the test or assessment are presented and on the vertical (x) axis, the frequency (or count) is presented. The frequency count for each score is plotted on the graph to give us our frequency distribution.

The Normal Distribution

Scores which, when plotted, form a smooth curve like the one depicted below are said to be ‘normally distributed’. This curve is sometimes called a ‘bell-shaped curve’. You can see that most scores fall around the average (bulge in the middle), with fewer occurrences towards the far left (low scores) and towards the right (high scores).

Most natural phenomena are normally distributed. If you were to plot the shoe sizes of a large number of women in the UK, you would discover that the distribution normal. You would find the same with



height and weight. Using these normal distributions, you are able to get a sense of where you stand compared to others. You can start to answer questions such as: Are my feet big, small or average? Am I tall, short or around normal? How well have I done in a numerical test? The normal distribution and its unique properties are the basis for all test norm systems.

Norm Systems

Grades/Bands

Grades or bands are a type of rank order norm system. The area under the curve represents the total percentage of people who have taken a particular test. We are able to chop up the curve into bands of average, above average, below average, high and low Or indeed grades of A, B, C, D and E. Grades and bands are one of the simplest norm systems we can use.

Benefits of using grades or bands:

Grades and bands are easy to understand and to interpret and hence they are often used in education to describe the performance of students.

Drawbacks of using grades or bands:

Grades and bands, by their nature, are not precise measurements. There are ‘grey’ areas at the borders and a single mark may mean the difference between an average and above average score. To overcome this lack of precision and differentiation, another norm system can be used: Percentiles.

Percentiles

Percentiles are essentially an extension of the Grade System – they are 'graded grades' whereby instead of having just five bands, you have many bands, giving you a more sophisticated grading system. In fact, the percentile system splits the normal distribution into 100 bands, each representing 1% of the comparison group or population under the curve. Percentile is defined as 'per cent', or 'of a hundred'.

The normal distribution can, therefore, be thought of as being divided up into percentiles. A percentile rank indicates the percentage of the norm group a person's score comes above. A score at the 60th percentile means that the individual performed better than 60% of the group (while 40% of the group have performed better than them). The way to describe a score at the 60th percentile is to say "you have performed better than 60 percent of the comparison group". This phrasing is useful when feeding back test scores to candidates or line managers.

Benefits of using percentiles:

Describing scores in terms of percentiles can be useful as they give a more precise measurement of a candidate's scores than grades or bands, enabling greater differentiation between candidates. Percentiles are commonly used with tests because they are easy to understand and to interpret.

Drawbacks of using percentiles:

However, percentiles are not equal units of measurement. An increase from the 95th to the 99th percentile is a greater performance improvement than an increase from the 50th to the 70th percentile. Percentile scores can therefore be said to reduce the differences at the extremes and exaggerate scores around the middle of the distribution. When using percentiles, it is, therefore, key that you do not over-read small differences between applicants; one or two raw scores differences near the average can result in a large gap in percentile terms.

This leads to a major practical problem - you cannot take percentiles from different tests to produce overall composite scores. In order to get around this problem, we use scores called standard scores, should an occasion arise when you wish to add or take an average of a set of scores. We will look at standard scores in detail later in the course.

Norm Groups

A norm group is the sample group against which a candidate's scores are compared. A norm group can be regarded as a sample, from which a set of scores have been gathered to provide a representation of the population it is intended to represent (e.g. managers, graduates, call center staff or the general population).

Norm groups should be up to date and in order to have statistical significance, should be based on a group of 100 + people. The extent to which the sample is representative of the population depends greatly on the size of the sample. Larger samples are more likely to be representative. For example, a norm group of a thousand Managers is more likely to be a reasonable representation of all Managers than a group of 50. The standard error of the mean (SEmean) allows us to estimate the distance of our sample mean from the population mean. It can be calculated using the mean of the sample, the SD and the sample size. As the sample size gets larger, the SEmean will reduce.

A large group, however, is not guaranteed to be representative and attention should be paid to the sampling method that has been used and whether there is an appropriate spread of people in the group. There are a number of different ways of collecting samples and developing norms. The main sampling methods include random, stratified and usage sampling.

In random sampling, you randomly select people to include in your sample, typically with the entire population having equal chance of being selected – this can be challenging to achieve when your population is large and broad, but can reduce sampling bias.

Stratified sampling is where you purposely select a sample that is representative of your intended population (e.g. if your population consists of 50% males and 50% females, you would ensure the same proportions in your sample).

However, even with a sophisticated methodology, there is still the challenge in assessment of the motivations of the candidates. Therefore, the final data from random or stratified samples are likely to be different from realistic live usage samples.

In usage sampling, you include those in your sample who have previously completed a test in a real application. Usage samples are often used to collect norm samples; an advantage of this is that those within the sample are usually realistically motivated to complete the test(s). Many usage samples and therefore norms are merely opportunist in nature, however, this is not the Saville approach.

We combine the advantages of using motivated usage data with the techniques of careful stratified sampling to create norm groups which are both representative and based on realistic test completion data. For example, Saville prevent over-sampling of particular groups when creating realistic usage norms by reducing the number of individuals selected from a particular country or client. This leads to smaller but better norm groups. Because we know that, statistically, beyond a few hundred people, sample size has relatively less impact than sample composition on norm quality, our best practice recommendation is to use realistic usage samples which are cut down to a smaller, more representative norm group.

Choosing Norm Groups

The group used to establish an appropriate norm group should always take into account the job being applied for. It would be appropriate to use a norm group of graduates for graduates entering the organization. It would not be appropriate to compare graduates' scores on a numerical test against a group of 16-year olds, nor would it be appropriate to compare them to a group of experienced Chartered Accountants. It is therefore important to take into consideration things like educational level and work experience in order to ensure that your norm group is representative of your candidate population.

Notes:

Module 5: Test Feedback

Purpose of Test Feedback

- For public relations – candidates are often clients too
- To assist in explaining why job offer not made
- To comply with applicable legislation
- To understand results by seeking examples/explanation
- To understand conflicts with other assessment data
- To increase recruited candidates' self awareness
- To develop individuals in key ability areas

Feedback of Saville Aptitude Tests

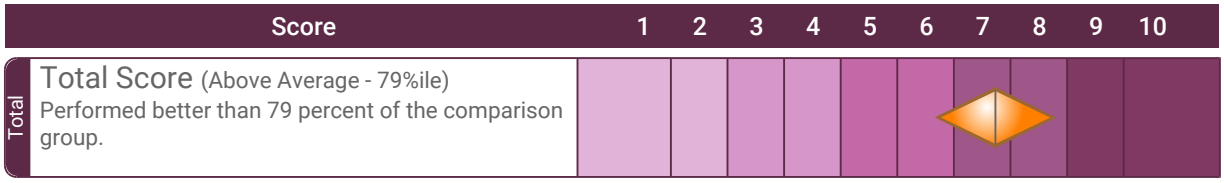
- **Total score** is main focus in feedback and decision making
- **Sub-scores** help to describe what the test is assessing
- **Item type** or **sub-scores** also help pinpoint very specific strengths and development needs
- Unsupervised online item-banked tests feature information about the combination of an individual's **Pace** and **Aptitude**
- Hard copy and supervised fixed-content tests contain information about an individual's **Speed**, **Accuracy** and **Caution**
- **Answer forms** and **feedback reports** give straightforward feedback and development advice for candidates, line managers and trained test users

Notes:

Unproctored Item-Banked Profile Example

Total Score

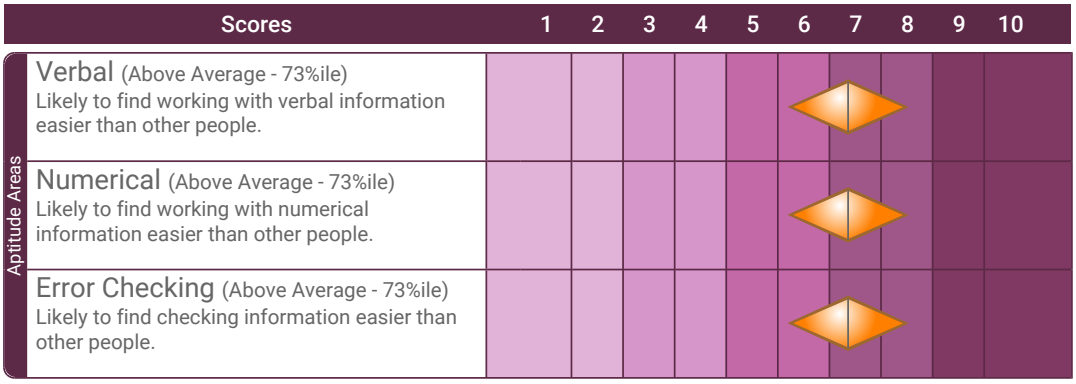
This page shows the Total Score relative to the Mixed Occupational Group (INT, IA, 2021) comparison group on a 1 to 10 sten scale.



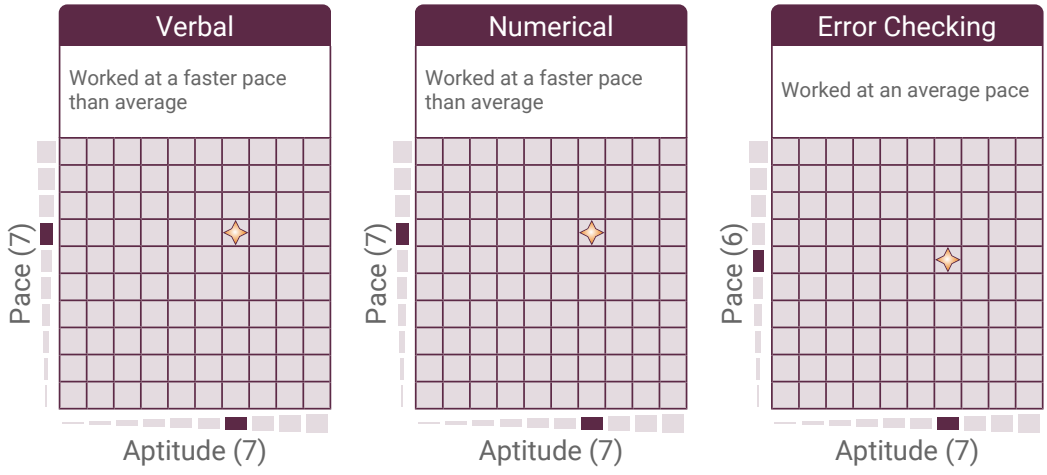
Aptitude & Pace Information

This page displays aptitude and pace information for each of the areas in the test relative to the Mixed Occupational Group (INT, IA, 2021) comparison group.

Aptitude Area Sub-Scores



Aptitude & Pace Comparison



Feedback Tips

More Effective Feedback

- Do build rapport (e.g. give eye contact, ask the delegate how they found the assessments)
- Do ensure two-way dialogue (e.g. invite the candidate to ask any questions throughout and be sure to check their understanding)
- Do gauge candidate reactions and impressions (e.g. ask the candidate whether their scores are in line with their expectations, ask what they think about their performance)
- Do discuss development areas if raised by candidate (e.g. if a candidate has referred to a particular area as a challenge for them, you may want to spend a little more time discussing relevant development tips)

Notes:

Less Effective Feedback

- Giving too much information or too many scores at once can confuse the candidate (i.e. give your candidate time to process the information and space to ask questions)
- Be mindful of using technical jargon (i.e. talking about specific report scores without explaining what they are or what they mean)
- Try not to make value judgments (i.e., “That’s really good…Oh that’s not so great.”)

Notes:

Aptitude Feedback Guide

Introduction

Self, purpose – yours/theirs, timings, confidential, data storage, two-way session.

Tests completed

Reminder, avoid jargon, why used tests, link to job role, how long results valid. Discusses test validity in layman's terms.

Invite questions

Checks understanding.

Candidate's experience

How found test session, self-evaluation– test like/ disliked, their background.

Comparison group

How scored, composition. You can use the About this Report section of the report to support you.

Total score

Percentiles/behavioral terms. Use this formula:

"You performed at the Xth percentile. This means that you performed better than X percent of the comparison group. This is a below average/ average/ above average score. What are your thoughts on this?"

Sub-scores

Percentile/behavioral terms. Break down in the same way as the Total Score, using this formula:

"You performed at the Xth percentile. This means that you performed better than X percent of the comparison group. This is a below average/ average/ above average score. What are your thoughts on this?"

Check in with the candidate between each sub score so that you're not overwhelming them with information.

Aptitude and pace

Describes Pace using non-technical terminology. Read the report verbalizers for pace straight from the report.

Candidate's response

Asks for reaction. Keep the candidate involved in the conversation, ask what they think after each section and invite questions.

Use of jargon

Appropriate language used.

Remain objective

Non-judgmental language, ask vs. tell. Only use the report language, use the behavioral terms, "below average/ average/ above average". Be careful about adding judgments, don't comment on how "good" a score is, just report it as it as on the profile.

Invites questions/checks understanding

Open questions, sandwiching your feedback with inviting questions.

Links results to other relevant information

Summary and conclusions.

General style

Build rapport, use appropriate style, empathetic but to the point.

Notes

Purpose of Test Feedback

Even if candidates are not successful, research shows that they are more positive about an organization if they have received feedback on their performance. Feedback also gives candidates a chance to understand why tests are used by the organization and the rationale behind using them. Candidates and feedback providers can discuss the pattern of results displayed, and discussion of examples from their working life can aid understanding. Where there are differences between ability test results and findings gained from other assessment center exercises involving numerical calculations, for example an in-tray forecasting exercise, these can be explored. Candidates can gain valuable insight into their relative strengths and development areas, which can help guide personal development in specific areas.

Candidates have the right to see any information held on them, including assessment results. It is best practice to provide this in an appropriate and accessible form, such as verbal or written feedback.

Feedback of Saville Assessment Tests

Saville Assessment ability tests give additional, unique information regarding test performance on different categories of questions on a test. These measures break down the overall total score into Item Type or Aptitude Area Sub-Scores, providing recruiters and candidates with an in-depth understanding of result patterns. The test-taking style measures provide added insight into how the candidate completed the test.

Pace – how quickly the individual has responded to the questions

Aptitude – how well the individual has performed on the test

Saville Assessment hard copy and online profile charts and feedback reports are designed to be used by a wide audience, including candidates, trained users and line managers, giving more flexibility. Graphic displays are used to ensure quick and straightforward interpretation and feedback.

Typical Feedback Structure

This structure can be used in a variety of contexts, e.g. face-to-face and telephone feedback.

The introduction would include you introducing yourself and your role within the assessment process. It is useful to establish the objective of the discussion, for example to provide feedback after the recruitment process. You could ask the candidate what specifically they would like to get out of the conversation. Ensure you include information such as how long the session will last and who else will see the test results.

It is worth reminding the candidate of the tests taken, using short and non-technical descriptions, e.g. ‘test designed to assess your ability to work with numbers’ for a numerical test. Mention the comparison group that candidates’ scores are being compared to, so they understand who their scores have been benchmarked against. It is also very useful to ask candidates how they found the tests as it can be useful for you to understand which ones they found easier and which more challenging. This can aid in the discussion of results later on.

Taking each test one by one, describe the candidate’s overall score on a test in behavioral terms (e.g. average) and also the proportion of the comparison group they did better than (percentile). Discuss patterns of test-taking style, e.g. slow but accurate and review any differences/patterns in the sub-scores on the profile charts.

Notes:

Module 6: Best Practice & Ethics

Best Practice: Key Points

- Promote proper data management
- Ensure fair assessment
  - Test validity
  - Preparation materials
  - Monitoring differences
- Consider the candidate's needs
- Make accommodations for special requirements
- Review your testing policy

Data Implications for Testing

General Data Protection Regulation (GDPR)

Applicable when assessing for candidates and/or clients in the EU.

Six principles to ensure that information is handled properly:

1. Processed lawfully, fairly and in a transparent manner
2. Collected for specified, explicit and legitimate purposes and not further processed for an incompatible purpose
3. Adequate, relevant and limited to what is necessary in relation to the purpose
4. Accurate and, where necessary, kept up-to-date
5. Kept in a form which permits identification of data subjects for no longer than is necessary for the purpose
6. Processed in a manner that ensures appropriate security of the personal data

Adverse Impact in Testing

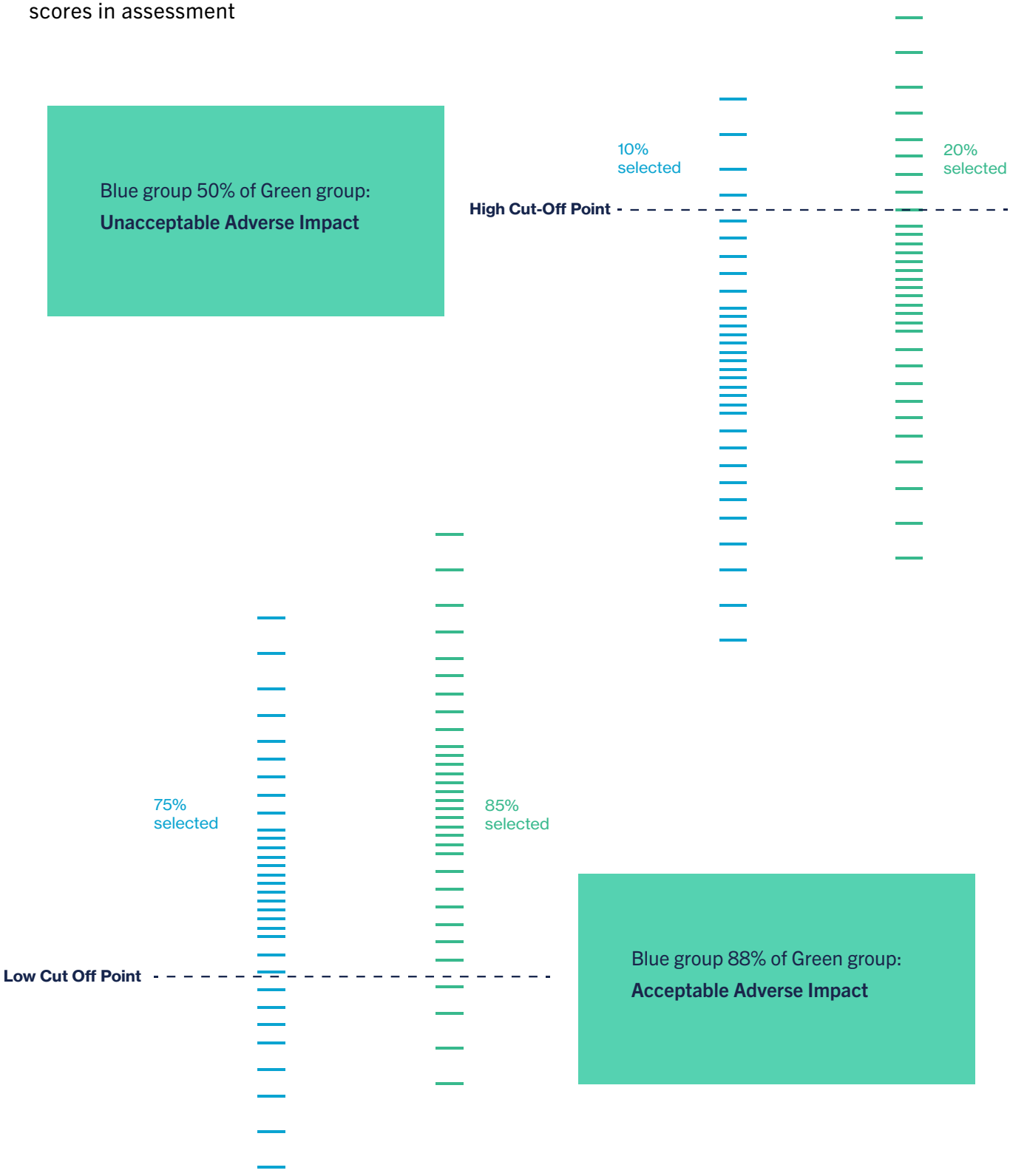
- Adverse impact is where a substantially different proportion of one group is hired in comparison to another group.
  - This is likely to occur where there is a difference between groups in terms of their average test score.
- As group differences exist, justification for using tests and the level of cut-off matters.
  - A high cut-off of, for example, the 90th percentile is likely to reduce the relative percentage of black candidates compared to white or Chinese candidates and it is likely to require strong justification to show that this level of aptitude is required.

Notes:

The Four-Fifths Rule

Purpose of Test Feedback

- A selection rate for any protected group which is less than four-fifths (or 80%) of the rate for the group with the highest rate of selection is seen as adverse impact
- This affects where we need to set cut-off scores in assessment



Best Practice: Consider the Candidate

- Explain the assessment process
  - Important for procedural justice
- Provide practice/preparation material
- Establish informed consent
- Provide opportunities for right of reply
- Establish feedback process
- Promote consistency and fairness for all candidates throughout the process
- Test in the most appropriate language (normally first language)
- Enquire about special requirements/accommodations (e.g. disabilities)

Testing Candidates with Additional Requirements

- You are responsible for deciding upon an appropriate accommodation/adjustment and will need to:
  - Understand the disability
  - Seek test publisher advice as to what different accommodation options are available
  - Seek expert advice on a candidate's disability where appropriate
  - Decide how to accommodate, test as standard, or not test at all
- Typical accommodations include: time adjustment, screen reader technology, large screen/print version, braille, response assistance (sighted reader or person to record responses)
- Conversations may be required with the disabled candidate, the assessment provider and even a suitably qualified professional for that disability
- Volume testers will generally have policies in place regarding common accommodations, e.g. particular time adjustments for dyslexia upon presentation of an appropriate certificate
- A larger band of error should be applied to adjusted test scores

Best Practice and Ethics

Proper Data Management – GDPR

When using assessments, you need to follow these six principle of the General Data Protection Regulation (GDPR).

1. Processed lawfully, fairly and in a transparent manner. The scores should be used to make fair decisions about people. This requires the use of well chosen tests with appropriate interpretation. Ensure that candidates are provided with sufficient information about the assessment process.
2. Collected for specified, explicit and legitimate purposes and not further processed for an incompatible purpose. Ensure scores are only used for the purposes for which they were collected. To use them for other purposes requires gaining further permission from the candidate. If an assessment is completed as part of a development process it is unlikely it would be appropriate to use the results for selection or promotion decisions at another time.
3. Adequate, relevant and limited to what is necessary in relation to the purpose. Ensure only appropriate tests are used. Tests are not used unless the information is needed for a proper business purpose, e.g. making effective selection decisions, developing staff.
4. Accurate and, where necessary, kept up-to-date. Ensure that care is taken in collecting and processing data to ensure it is accurate.
5. Kept in a form which permits identification of data subjects for no longer than is necessary for the purpose. That there is a policy of deleting data once it is no longer useful. Typically test scores remain relevant for 12-24 months. After this they should be erased.
6. Processed in a manner that ensures appropriate security of the personal data. Appropriate security should be in place when storing data. Appropriate technical or organizational measures should be in place to protect against unauthorized or unlawful processing and against accidental loss, destruction or damage. Each organization should take their own legal advice with regard to their human resource activities. Saville Assessment is not in a position to advise on legal matters.

Equal Opportunities Legislation

Equal opportunities legislation has developed over time to protect more groups, with major legislative developments in the latter half of the 20th Century. This legislation has continued to strengthen and evolve to cover more protected groups.

For example, the UK Equality Act 2010 protects the following characteristics:

- age
- disability
- gender reassignment
- marriage and civil partnership
- pregnancy and maternity
- race
- religion or belief
- sex
- sexual orientation

Discrimination

Unfair treatment of any of these groups would be considered as discrimination. Discrimination may be Indirect or Direct.

Indirect Discrimination

Indirect Discrimination is the unintentional differential treatment or adverse impact that affects different groups as a result of the testing conditions imposed. Hiring managers should consider whether there is clear justification for their testing choice, for example, it would be indirect discrimination to ask one group of candidates to complete an English language test but not asking all of their candidates to do this.

- The unintentional differential treatment of candidates in different groups
- Testing decisions need to be justifiable if it could be claimed that indirect discrimination has occurred, for instance, the cut-score in a selection process negatively impacts a particular group but it is vital for selected candidates to have that level of performance in a given area
- Be sure to select tests that have minimal observed group differences

Direct Discrimination

Direct Discrimination treats people differently because of the group they belong to; this is almost universally outlawed and this is not something that any high-quality assessment is designed to do. An example of direct discrimination of assessment could be not allowing a person with a disability to complete a test as part of a selection process.

- The intentional differential treatment of people depending on a certain group they may be part of, such as gender, race or religion
- High-quality assessments are not designed to be used in this way

Fair Assessment

Respect for the Individual

It can be beneficial for the administrator to understand candidates' concerns or overall perspective of the experience. Ensure that candidates are fully briefed on the rationale and processes used to reach decisions and that you treat candidates with understanding.

Administrators need to deal with questions and problems in a patient and professional manner. Testing may be unfamiliar to candidates and they may be surprised by the formal nature of administration. There is evidence that candidates are more likely to regard decisions as fair when they are aware of the processes used to reach the decision.

- Administrators should treat candidates as they would like to be treated in the process
- Be sure that the candidates are aware of the process and why it is being used
- Demonstrate understanding of the nerves a candidate may experience

Choosing Appropriate Tests/ Questionnaires

As we learned in Job Analysis and Assessment Choice, tests and questionnaires should be chosen on the basis of a thorough job analysis to ensure that decisions are being made on the basis of relevant information. To ensure assessment fairness, look for evidence of studies examining the appropriateness of the instrument with different groups.

- Assessment choice should be based upon thorough job analysis
- Consider reviewing technical summaries for evidence regarding the appropriateness of test use with different groups, e.g. validation studies

Preparing the Candidate

This is particularly important for aptitude tests. It is usually recommended to advise candidates how their data will be used, how they will be stored and whether they may be used again in the future. Candidates can access practice tests and guidance on the Saville Assessment website,and are included on the candidate dashboard.

Candidates should be briefed ahead of completing a psychometric test:

- The rationale for using the assessment
- What the assessment measure
- How their data will be stored and who will have access to their results
- Whether they require any reasonable adjustments
- Gaining informed consent from the candidate

Dealing with Language

For any psychometric measure you should consider what the impact of language needs are. Where English is not the primary language, consider whether it would be more appropriate to test in another language. Where an organization considers English to be the required language they may feel that testing candidates in English is justified. However, it is generally recommended that candidates are tested in their preferred language where possible. English proficiency assessments are available alongside aptitude assessments.

It is generally recommended to assess candidates in their language of greatest proficiency, wherever possible.

If you are not sure of the implications for testing, you can contact us.

Disability Considerations

Disability adjustments should be managed on a case by case basis, discuss any issues with candidates ahead of assessment to understand and accommodate their needs.

Some examples include:

- A candidate with dyslexia may have difficulty reading some assessment content and may need more time than other candidates to complete the task
- A candidate with sight impairment may have difficulty reading a booklet or seeing a computer screen; the candidate may need to use screen-reader software or have assistance from a sighted facilitator
- A candidate with a motor impairment may have

difficulty using a mouse to fill in an answer sheet, so could instead use touch-screen functionality

- Manage candidate needs on a case-by-case basis
- Ask the candidate to provide what has been recommended for them by an appropriate specialist, e.g. An educational psychologist has recommended additional time for a person with dyslexia
- The general principle is that any adjustment should attempt to provide the individual with a comparable assessment experience to other candidates
- The assessor and assessee both have a responsibility for being as accommodating as is reasonable
- Saville Assessment online aptitude assessments are compatible with all modern computer and tablet browsers which permits a range of adjustments
- Where you are unsure of the appropriate reasonable adjustments to make, you should seek expert advice

Using Tests Responsibly

Interpreting Score

Care should always be taken to interpret an assessment correctly, being clear on what the different aptitude areas measure and what scores mean. You can use the assessment descriptions in the technical manuals to support you. Consider the appropriate scales to feedback to candidates, the most suitable comparison groups and whether any reasonable adjustments made have impacted test scores. Remember to take into account the size of error around their score and how they perform in comparison to the benchmark group.

- Make sure you know what the assessments you are using are measuring
- Be clear on how to interpret scores, their error of measurement and how best to give feedback on these to a candidate

Feedback

Candidates are likely to be interested in their results. Giving the option to have written or spoken feedback is recommended and in some regions, candidates have a legal right to access their results. This can help to increase candidates' self awareness and better understand how their results have been used in the decision-making process. This is likely to make candidates feel more comfortable about the way in which their results are used in selection and development processes.

- Feedback may be a legal requirement based on the country in which the process takes place
- Feedback can help the candidate's self-awareness and understanding of the process

Test Use Policy

It is generally good practice for the use of tests to be guided by a test use policy. This will set out standards and local policies on a range of relevant issues. This helps ensure that minimum standards are maintained and that there is a consistency in practice across different assessment processes.

- Your organization should have and use a test use policy
- A test policy outlines the standards and requirements to be used consistently through your organization's testing processes
- A sample test use policy is available from us

Training and Responsibility of Test Users

It is important to complete training before using some assessments but, as with any skills or knowledge, over time parts may be forgotten and bad habits can develop. Equally, new developments may require updating of knowledge. Engaging with these developments to maintain up-to-date knowledge and develop skills means that you can continue making best use of assessments. It is the responsibility of the test administrator to ensure proper practice and ensure that all interpretations from the test are valid and appropriate to the context and for the person who is using the information.

- It is important to complete appropriate training ahead of using some assessments
- Test administrators should stay up to date with any new developments to ensure they are delivering best practice assessment use

Best Practice: Key Points

When using aptitude tests it is important to consider points of best practice:

- Promote proper data management
- Ensure fair assessment
- Use valid tests
- Provide preparation materials to candidates
- Monitor group differences in the samples you progress at each assessment stage
- Consider candidates' needs
- Make accommodations for special requirements
- Review your testing policyes

Module 7: Reliability & Validity

Reliability

Reliability is fundamental to measurement and concerns how precise and error-free a tool is in measuring desired constructs. Any instrument that measures something in the real world needs to have a level of precision or accuracy, for example, weighing scales, a digital clock or a light meter in a camera. The greater the reliability or precision, the greater the chance that it will allow for valid decision-making.

- Reliability is concerned with how precise and error-free a tool is in measuring intended constructs
- Any instruments of measurement need to have a level of reliability, or precision, to be useful
- Regarding behavioral measures, the greater the reliability, the greater the chance of making a valid testing decision in selection or development

Types of Reliability

Test-Retest

Test-retest reliability refers to the stability of a measure over time. It is calculated by correlating results from a measure completed by the same group of people at two points in time.

+ Gives indication that attribute is stable

- Candidates not willing to do it twice

Alternate or Parallel Form Reliability

Alternate or Parallel form reliability refers to the consistency between two versions of the same measure. This is the correlation between the results for the same group of people who complete two versions of the questionnaire.

+ Shows developer is clear/consistent on what is measured

- Has the expense of developing two forms

Internal Consistency Reliability

Internal Consistency reliability relates to the internal correlations of the components of the measure, for example the relationship between the different scales within one questionnaire.

+ Easy to do as only requires one set of data from one time period

- Can be misleadingly high with repetitive item content

While all forms of reliability are important, internal consistency is often the most practical and accessible form of reliability, which can be more readily calculated in large samples. The generally accepted benchmark level for test reliability is r = +.70.

Error

Self-report scores can contain errors of measurement for a number of reasons.

Individual

If the individual feels unwell, has not given themselves appropriate time, misinterprets the questionnaire instructions or experiences severe test related anxiety, these factors could all mean they may not complete a questionnaire properly.

- Feeling unwell
- Misinterpreting instructions
- Severe test anxiety

Administration

If the test administrator has chosen a test which doesn't accurately measure what it claims to measure, e.g. a behavioral measure with very little workplace validity, this can be a form of error. Likewise, when administrators do not properly brief candidates or set up the testing environment appropriately, to minimize disruptions for example, this results in distractions which can reduce a questionnaire's reliability. The administrator should diligently mark any hard-copy responses, where used, and be sure to accurately interpret results; where this is not the case assessment error is introduced and the reliability of the results will be lowered.

- Using an unreliable test
- Poor candidate briefing
- Misinterpreting responses

Questionnaire Developer

Questionnaire developers should be rigorous in ensuring the quality of their measures to support the reliability of their findings. This includes writing clear questions or items which lack any ambiguity, giving straight-forward instructions and being sure that their assessments are measuring what they claim to measure. Reliability is about getting the test right; validity is about getting the right test. It is the test developer's responsibility to develop an accurate test and ensure it is a reliable measure.

- Ambiguous items
- Items measuring the wrong thing
- Poor instructions

An example of an ambiguous item could be one that uses a colloquialism or metaphor such as, 'I often feel blue'. This may not translate well into a number of languages and could be confusing to individuals completing the questionnaire.

Reliability and Error

Scores obtained in occupational questionnaires invariably contain a degree of error. The Standard Error of Measurement, or SEm, takes this error into account when dealing with individual responses. That is, the SEm measures the margin for error in an individual's score. It enables us to assess the confidence we can have in the precision of an individual's score, by presenting a band in which we are confident their score lies. When a score lies in a band of plus or minus one SEm, we have a 68% confidence level in the score being accurate. A band of two SEms reflects a confidence of 96% accuracy. The use of the SEm means that scores can be generalized across the population, using confidence levels. The typical SEm of Wave Professional Styles is slightly less than one Sten. This means an individual's true measure is likely to be within one sten score of what is reported on their Wave profile.

- All behavior tools have a degree of error
- Standard Error of Measurement (SEm) accounts for this error
- SEm provides a band in which we are confident that an individual's true score lies
- The typical SEm of Wave Professional Styles is slightly less than one Sten, this means that an individual's true response is likely to always be within around one Sten of what is shown on their profile

Validity

A test is valid to the extent that it measures what it is designed to measure. In particular, validity is a measure of how relevant a behavioral questionnaire is to job content. This is a key aspect of using occupational tools; if the tool is not valid, then there is little point in using it. You may have a highly reliable questionnaire, but if it is not measuring the particular job skills potential you are interested in assessing, then it is not useful. Remember, that a valid tool has to be reliable in the first place. Studies generally indicate that a good personality questionnaire can have a validity of +0.3. Validities above +0.7 are virtually unknown in the literature. The higher the validity, the better.

Types of Validity

Assessment validity can be thought of as Informal or Formal. Informal types of validity are more concerned with how a test appears whereas Formal types of validity are more rigorous.

Informal

Face Validity

Tools with high face validity ensure buy-in from candidates and line managers, but with face validity alone, questionnaire choice is not based on hard evidence and is unlikely to be legally defensible if challenged. However, it may be the lack of face validity which instigates a legal challenge when candidates question the relevance of the questions they are being asked in relation to performing effectively on the job.

Face validity looks at whether the instrument appears to be measuring what it should be. Questionnaire items should be written with face validity in mind to ensure that the questionnaire 'looks right' and that it is acceptable to individuals completing it. It is important to remember that whilst face-validity is important for buy-in from candidates and users it does not guarantee any statistical robustness of the tool. Using tools that lack psychometric robustness can lead to mistakes in selection & development, and feedback & interpretation.

Faith Validity

Faith validity is a spurious form of validity. It is an unquestioning belief that a questionnaire is appropriate and predictive of job effectiveness. Faith validity can aid in getting buy-in for the use of objective assessment methods. However, lacking hard evidence of robust assessments can lead to misuse of tools and in the worst case scenario could lead to the use of measures that are not legally defensible or valid, which don't allow for the selection of better candidates.

An unfounded belief that a tool is appropriate and effective; a feeling that the test works in the absence of evidence. Faith validity is the least defensible form of validity.

Barnum Effect

A 'Barnum effect' occurs when a statement in a questionnaire, or a description on a profile, is phrased in such a way that it could be applicable to anyone.

Consequently, a candidate's positive response to such a statement has minimal value since all candidates are likely to agree with this statement.

- The phrasing of questionnaire statements or profile descriptions mean that they could be applicable to anyone
- Responses to such items have minimal value as most candidates will respond similarly

Formal

Consequential Validity

The intended and unintended consequences of using a test. Test users should be mindful of how their use of assessments could impact assesses. For example, when using assessments to identify high potential there is the intended consequence of encouraging individuals to develop in relevant areas. An unintended consequence could be narrowing individuals' focus to just those areas being assessed rather than other relevant work areas.

Content Validity

Content validity reflects the extent to which the items in an instrument are representative of job-relevant content. Wave Professional Styles has been designed to measure a core set of personality characteristics required for a broad range of roles. The items cover both the Talent (e.g. 'I am good at selling') and Motive (e.g. 'I enjoy selling') aspects of the personality dimensions being measured. In the development of Wave, a research and conceptually-driven hierarchical model was created, which maps to the Wave skills potential framework. Items were written and refined based on statistical analyzes and professional expertise.

- Content validity refers to the relevance of the items of an instrument to job-related content
- Wave Styles questionnaires measure core personality characteristics relevant to a number of roles
- Wave Styles capture both self-perceived Motive and Talent related to such areas
- Research and a conceptually-driven approach led to the development of the Wave Styles and Skills Potential frameworks
- Wave items were written and refined based on statistical analyzes and professional expertise

Construct Validity

Construct validity concerns the extent to which an instrument measures some underlying theoretical construct or trait. Wave Styles has been designed capture the 'Big Five' model, as well as skills potential constructs such as the 'Great Eight' model. At the same time, we retained important work constructs even if they did not fit neatly into established academic theories.

- Construct validity pertains to the extent to which an instrument measures an underlying theoretical construct or trait
- Wave Styles was developed to capture the Big Five personality theory and Great Eight Work Performance model

Criterion-related Validity

Criterion-related validity is the extent to which a questionnaire is able to predict job performance variables such as appraisal ratings, potential for promotion and achievement of targets and objectives. The most common way of establishing criterion-related validity is by correlating questionnaire scores with measures of job performance. The main methods of approach to this are through concurrent validation and predictive validation.

Refers to evidence that the test predicts relevant criteria (e.g. skills potential or workplace outcomes).

Concurrent

The potential effectiveness of a new questionnaire is investigated on current employees within an organization.

Predictive

The impact of a new questionnaire is evaluated by following up the performance of selected individuals some months after being recruited.

Validity? So What?

- Questionnaires with the highest validity increase the chance of selecting the best performers at work and considerably reduce selection errors
- Wave Styles was developed to capture the Big Five personality theory and Great Eight Work Performance model
  - This can also reduce the number of serious selection errors five-fold

0 Validity - 1 person in 5 will be a poor performer

0.3 Validity - 1 person in 10 will be a poor performer

0.6 Validity - 1 person in 50 will be a poor performer

Notes

Reliability

Definition

Reliability is fundamental to measurement and concerns how precise and error-free a tool is in measuring desired constructs. In self-report questionnaires, reliability concerns how consistently and precisely a questionnaire measures a characteristic. Reliability is important when interpreting personality assessment scores, because they are intended to reflect the individual's true personality. Reliability is crucial for validity, as an inconsistent or unreliable measure cannot be valid because its lack of reliability restricts the true measurement of personality.

Types of Reliability

Test-Retest reliability refers to the stability of a measure over time. It is calculated by correlating scores on a measure completed by the same group of people at two points in time.

Alternate or Parallel Form Reliability refers to the consistency between two versions of the same measure. This is the correlation between the results for the same group of people who complete two versions of the questionnaire.

Internal Consistency Reliability relates to the internal correlations of the components of the measure, for example the relationship between the different scales within an assessment.

For self-report questionnaires it is important that internal consistency reliability is satisfactorily high but not artificially inflated. Narrow scales with repetitive item content have high reliability but lack breadth of measurement. In the development of Wave Professional Styles this problem was avoided by drawing on three distinct facet constructs for each dimension.

Wave Professional Styles Reliability

A development goal of the Wave Styles assessments was to have alternate form and test-retest reliabilities as high as possible. The Wave Styles assessments were designed to have moderate (0.6 – 0.9), rather than high internal consistencies at the dimension level (as they are made up of six different work constructs – motive and talent).

The 36 dimensions of Wave Professional Styles demonstrate acceptable test-retest reliabilities over an 18-month interval with coefficients ranging from .58 (Principled) to .85 (Activity Oriented) and a mean reliability coefficient of .75.

The alternate form reliability of Saville Assessment Wave Professional Styles is based on two versions of Professional Styles; Invited Access (IA) and Proctored Access (SA). At the dimension level, the mean reliability of the dimension scales (combined Normative and Ipsative) was .86 and the minimum reliability estimate for any dimension was .78.

The dimensions of Wave Professional Styles were designed to have internal consistency estimates ranging from .60 to a maximum of .90. The mean internal consistency is in the center of this desired range, at .74. Only one scale fell outside this – Insightful, with an internal consistency of .58. However, Insightful has highly acceptable alternate form reliability and test-retest reliability estimates which are the fundamental reliability measures for Wave Styles.

Error

Sources of Error

Self-report scores can contain errors of measurement for a number of reasons:

Individual - The individual completing the assessment may have been feeling unwell on the day or may have had a 'bad' day, both of which can influence an individual's responses. The reasons for completing a questionnaire can also impact on responses; for example, if completing a personality measure as part of a selection procedure, the individual's perception of the organization's values may bias their responses. The environment can also impact on the reliability of assessment scores. The conditions (heat, noise levels) in which individuals complete the assessment can also influence response style.

Administration - The way in which the assessment is administered is also crucial to the reliability. As Wave Styles is an online measure, it is more immune to these sources of error, however, administrators should ensure a clear rationale for using the assessment is provided.

Test Developer – The construction of an assessment can impact on its reliability. For example, if questions are ambiguous or don't measure the intended construct the assessment is less likely to be reliable.

Reliability and Error

Scores obtained in occupational questionnaires invariably contain a degree of error. The Standard Error of Measurement (SEm) enables us to make allowance for this error when dealing with individual scores. Thus the SEm is concerned with the margin for error in an individual's score. It can, therefore, be used to assess the confidence we can have in the precision of an individual's score, by presenting a band in which we are confident the individual's score lies.

When a score lies in a band of plus or minus one SEm, we have a 68% confidence level in the score being accurate. A band of two SEms reflects a confidence of 96% accuracy. The use of the SEm means that scores can be generalized across the population, using confidence levels. The typical SEm of Wave Professional Styles is slightly less than one Sten.

Validity

What is Validity?

A questionnaire is valid to the extent that it measures what it is designed to measure. In particular, validity is a measure of how relevant a questionnaire is to job content. In developing Wave Styles, providing validation evidence was considered paramount in presenting a questionnaire that is based on a robust model of personality and is relevant to the workplace. As such, validation was incorporated into the construction of Wave Styles from its inception.

Types of Validity

Face validity looks at whether the instrument appears to be measuring what it should be. In the construction of Saville Assessment Wave, great care was taken to avoid items that lack face validity in a work context such as questions related to neuroticism and clinical symptoms. Although face validity has no statistical basis, it is essential that a questionnaire 'looks right'; that is, it appears to measure what it is intended to measure, for example, personality characteristics required in the workplace. A questionnaire that is face valid is one that it is acceptable not only to the individuals who complete it but also to those who will be required to interpret and act upon its findings. Members of an organization are more likely to feel comfortable in their use of a questionnaire and individuals more readily accepting of the results if the questionnaire appears reasonable and appropriate to them.

However, there is a danger that users may rely on spurious validity, such as face validity, as evidence of its true validity. It cannot be assumed, for example, that because a questionnaire is face valid, that it is also psychometrically valid. Using a questionnaire that is not psychometrically robust can subsequently lead to mistakes in selection, development, feedback and interpretation.

Content validity reflects the extent to which the items in an instrument are representative of job-relevant content. Wave Professional Styles has been designed to measure a core set of personality characteristics required for a range of professional and managerial roles. The items cover both the Talent (e.g. 'I am good at selling') and Motive (e.g. 'I enjoy selling') aspects of the personality dimensions being measured. In the development of Wave, a research- and conceptually driven hierarchical model was created, which maps to the Wave skills potential framework. Items were written and refined based on statistical analyzes and professional expertise.

Consequential validity considers the intended and unintended consequences of using a questionnaire. For example, if an assessment is being used to identify high potential people within an organization for succession planning purposes, intended consequences could include encouraging individuals to strive to develop themselves in performance-relevant areas, greater motivation and effort displayed by potential succession candidates and improved understanding of what matters for effective performance. On the other hand, unintended consequences could include a narrowing of focus amongst potential succession candidates to just those variables assessed by the questionnaire, potential succession candidates engaging in practices to disadvantage others and inappropriate use of assessment scores by the administrators or decision-makers.

Construct validity concerns the extent to which an instrument measures some underlying theoretical construct or trait. Professional Styles has been designed to comfortably cover the scope of leading personality theories such as the 'Big 5' model, as well as constructs such as the 'Great 8' model. At the same time, we retained important work constructs even if they did not fit neatly into 'parsimonious' academic theories.

Faith validity is a spurious form of validity. Faith validity is a blind belief that a questionnaire is appropriate and predictive of job effectiveness, for example, because of the plausibility of scale names or the acceptability of the report by candidates. A 'Barnum effect' occurs when a statement in a questionnaire is phrased in such a way that it could be applicable to anyone. Consequently, a candidate's positive response to such a statement has minimal value since all candidates are likely to agree with this statement. Faith validity is the least defensible form of validity.

Finally, **criterion-related validity** is the extent to which a questionnaire is able to predict job performance variables such as appraisal ratings, potential for promotion and achievement of targets and objectives. The most common way of establishing criterion-related validity is by correlating questionnaire scores with measures of job performance. The main methods of approach to this are through concurrent validation and predictive validation.

- **Concurrent validity** - the potential effectiveness of a new questionnaire is investigated on current employees within an organization.
- **Predictive validity** – the impact of a new questionnaire is evaluated by following up the performance of selected individuals some months after being recruited.

Increasing Validity Increases Return on Investment

Questionnaires with the highest validity increase the chance of selecting the best performers at work and considerably reduce selection errors.

An example of a serious selection error is selecting a candidate from the bottom 20% of performers when you mean to select from the top 20% of performers:

(i) If a questionnaire has a validity of 0.0, one person in every five that you select will prove to be in the bottom 20% of performers.

(ii) If a questionnaire has a validity of +0.3, one person in every 10 that you select will prove to be in the bottom 20% of performers.

(iii) If a questionnaire has a validity of +0.6, one person in every 50 that you select will prove to be in the bottom 20% of performers.

Moving from recruitment using a questionnaire with a validity of +0.3 to using a questionnaire with a validity of +0.6 can double the cost-benefit to an organization. It can reduce the number of serious selection errors five-fold, remarkably improving the accuracy of the selection process.

Saville
assessment

info@savilleassessment.com
www.savilleassessment.com