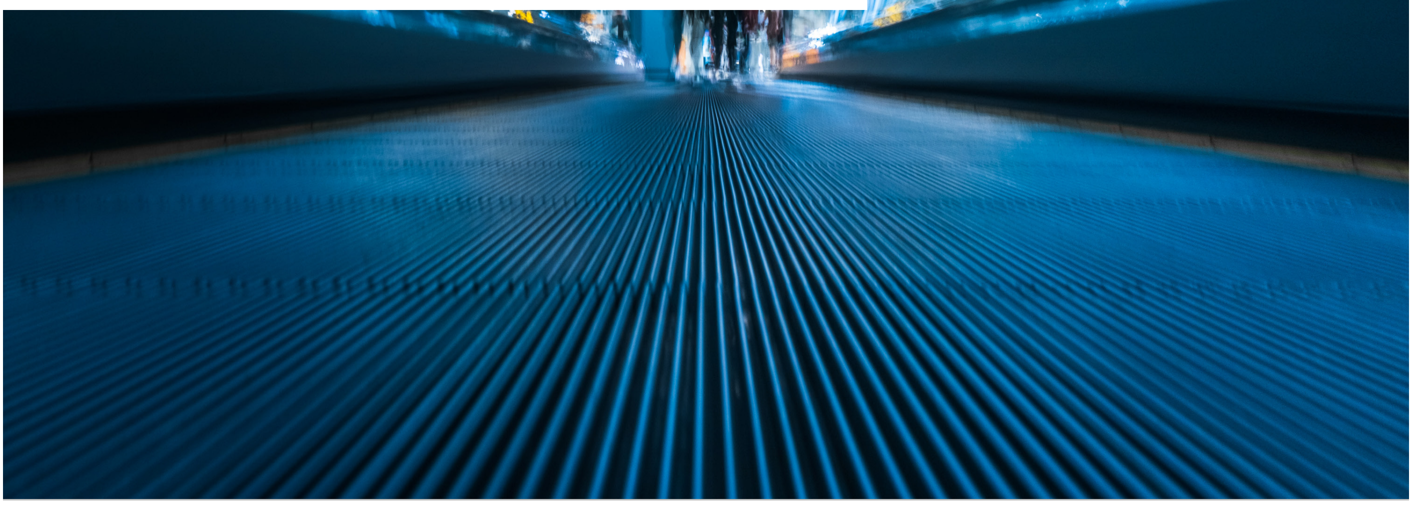
Technical Aptitude Range

Technical Summary



info@savilleassessment.com | +44 (0)20 8619 9000 | www.savilleassessment.com

Saville Assessment
A WTW Company

CONTENTS

1.0 Technical Aptitude Range Test Information 2

2.0 Norm Groups 2

3.0 Reports 2

4.0 Practice and Preparation 2

5.0 Development 3

6.0 Languages 3

7.0 Reliability 3

8.0 Validity 4

9.0 Fairness 5

10.0 Appendix 1: Internal Consistency Reliabilities for Sub-Tests in Swift Technical Aptitude 8

11.0 Appendix 2: Method for Calculating Criterion Related Validity of a Single Test from the Equivalent Sub-Test in Swift Technical Aptitude 8

Technical Aptitude Range Technical Summary

1.0 Technical Aptitude Range Test Information

Key Information

The tests in this range measure the ability to reason with information presented in different formats:

- **Spatial Reasoning** assesses the ability to visually rotate shapes, judge sizes and compare three-dimensional objects
- **Mechanical Reasoning** assesses the ability to comprehend mechanical problems, physical principles and movement of objects
- **Diagrammatic Reasoning** assesses the ability to analyze diagrams, sequences and transformations

Technical Information

- Technology supporting individual time limit for groups of four questions (testlets)
- Linear-on-the-fly (LOFT) testing
  - Fixed-length test
  - Based on Item Response Theory (IRT) methodology and scoring mechanism
  - Draws items of equivalent difficulty from a bank of items for different candidates
- Available for unsupervised use online (Invited Access, IA)
- Compatible with tablets, laptops and desktop computers

Test	Total / Sub-Test	No. of Questions	Time Limit (mins)
Swift Technical Aptitude	Total	28	10
	Spatial Reasoning	12	3
	Mechanical Reasoning	8	3
	Diagrammatic Reasoning	8	4
Spatial Reasoning Aptitude	Total	32	8
Mechanical Reasoning Aptitude	Total	32	12
Diagrammatic Reasoning Aptitude	Total	32	16

Note: Supervised Access (SA) Technical Aptitude Range tests using fixed content presented in a fixed order are available for follow-up testing but are not covered in this summary document.

2.0 Norm Groups

A range of international and country specific norms are available for the tests in this range. Information on the latest norm availability, norm group descriptions and other support documentation for norms can be found in the Client Area on the Saville Assessment website (www.savilleassessment.com).

3.0 Reports

Example reports for all the tests in this range can be found in the Client Area on the Saville Assessment website.

4.0 Practice and Preparation

Online practice tests are available for all the tests in this range. They are designed to provide a realistic set of example questions in order to help familiarize the test taker with the format and style of the aptitude assessment questions, as well as additional information about the assessment process.

The online practice tests also provide individual feedback on the responses given, featuring realistic time limits which replicate a real assessment scenario.

The aptitude practice and preparation materials can be found on the Saville Assessment website.

5.0 Development

The Technical Aptitude Range consists of large banks of Spatial Reasoning, Mechanical Reasoning and Diagrammatic Reasoning items. Items are drawn from these banks to form the single Spatial Reasoning, Mechanical Reasoning and Diagrammatic Reasoning tests and the corresponding sub-tests of Swift Technical Aptitude.

6.0 Languages

We are engaged in an ongoing, active program of translation and localization for all of our aptitude assessments. For the latest availability information, please contact Saville Assessment.

7.0 Reliability

The internal consistency figures presented here are Separation Indices. This method produces similar figures to Cronbach's Alpha (Andrich, 1982[1]) and allows for an internal consistency calculation to be made in item-banked tests, rather than fixed-form tests.

This section presents internal consistency reliability figures for each of the Technical Aptitude Range tests.

For Swift Technical Aptitude, it is worth noting that the greatest level of reliability is found at the total score level, which is designed to be the decision-making score. The sub-test scores provide additional test-taking information, but we would not recommend that these are used in isolation for decision making.

The mean percentage correct figures broadly reflect the design aim of giving a positive candidate experience where many candidates answer around 50% of questions correctly.

The large standard deviation values seen in these tables reflect the ability of the items to differentiate performance through a wide score range. This is required to give an accurate representation of test-takers' ability.

Swift Technical Aptitude (IA) Internal Consistency Reliabilities (N=10682)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	54.03	16.38	.95	4.74	**.77**

Spatial Reasoning Aptitude (IA) Internal Consistency Reliability (N=5430)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	53.15	16.90	.88	4.38	**.81**

Mechanical Reasoning Aptitude (IA) Internal Consistency Reliability (N=6492)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	55.39	17.35	.90	4.52	**.80**

Diagrammatic Reasoning Aptitude (IA) Internal Consistency Reliability (N=1581)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	65.33	16.42	.79	3.97	**.84**

[1] *Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. Education Research and Perspectives, 9(1), 95-104.*

8.0 Validity

Swift Technical Aptitude

This summary document includes criterion-related validity information for the total score and three sub-tests in Swift Technical Aptitude, based on a sample of 308 individuals for whom third-party ratings of workplace performance were collected. The criteria used here represent a priori predictions of the areas of work performance which each test was designed to predict.

The internal consistency of the summed criterion used is .79. This suggests that it is an acceptable assumption to combine the three separate workplace criteria to make a total criterion measure. Because N=263 of this sample of respondents also engaged a second rater of their workplace effectiveness, it was possible to take into account the inter-rater reliability of the criterion which can artificially limit the validity estimate. The inter-rater reliability measure takes into account the fact that there will always be some degree of difference between multiple raters' judgments of effectiveness on the criteria of interest, which can force the validity coefficient down.

The greatest validity contribution comes from the Mechanical Reasoning sub-test, with the least coming from the Spatial Reasoning sub-test. The trial Spatial Reasoning test version used in this sample was made up of just 8 items and has since been increased to 12 items. As a result, the validity of this sub-test is expected to increase, thus enhancing the overall validity figure.

For further information about the criterion-related and other forms of validity evidence for Technical Aptitude Range assessments, please contact Saville Assessment.

Swift Technical Aptitude Criterion-Related Validity (N=308)

	Correlation with Sum of Working with Designs, Equipment and Systems (Rater)	
	Uncorrected r	**Corrected r**
Total	.23	.41
	Correlation with Working with Designs (Rater)	
	Uncorrected r	**Corrected r**
Spatial	.06	.10
	Correlation with Working with Equipment (Rater)	
	Uncorrected r	**Corrected r**
Mechanical	.24	.39
	Correlation with Working with Systems (Rater)	
	Uncorrected r	**Corrected r**
Diagrammatic	.19	.45

Note: Any raw correlation higher than .12 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .10 is statistically significant at the p<.05 level (one-tailed). N=308. The criterion inter-rater reliability figures from Project Epsom (N=263) and the corrected figures are based on the inter-rater reliability figures for each of the Working with Designs, Equipment and Systems criteria (.33, .36, and .18 respectively). The criterion internal consistency of ratings (N=308) was .79 (Designs, Equipment, Systems; together known as Working with Things). Other than taking into account unreliability of the criterion measure, there has been no other adjustment for any statistical artefacts applied.

Technical Aptitude Range Single Tests

The Spatial, Mechanical and Diagrammatic Reasoning single tests are longer than the Swift Technical Aptitude combined assessment and cover the same areas of aptitude in greater depth. It is appropriate to assume that the Swift Technical Aptitude validities are a conservative and lower-bound estimate of the validity of the Technical Aptitude single tests, which are likely to show incremental validity over the Swift assessment (see Appendix 2).

9.0 Fairness

Gender Group Differences

Total Score - Swift Technical Aptitude

Test	Male N	Male Mean	Male SD	Female N	Female Mean	Female SD	Pooled SD Difference
STA Total	2948	-.15	.60	1819	-.11	.54	-.08

By Measure

Test	Test	Male N	Male Mean	Male SD	Female N	Female Mean	Female SD	Pooled SD Difference
Spatial*	STA sub-test	2948	-.18	.80	1819	-.04	.76	-.19
Mechanical*	STA sub-test	2948	-.12	.70	1819	-.32	.63	.29
Diagrammatic*	STA sub-test	2948	-.16	.83	1819	.03	.77	-.23

No comparison on the Technical Aptitude Range single tests due to insufficient data from the Female group (all N<500).

The tables above present the gender group differences on the Swift Technical Aptitude Total Score, and separately the sub-tests that measure each of the three aptitude areas in Technical range – spatial, mechanical and diagrammatic.

Expressed in terms of raw theta (ability) scores, there was no notable difference between men and women on the Swift Technical Aptitude Total Score. There was also no notable difference between the two gender groups on the spatial sub-test (less than .20 of a standard deviation). For the mechanical sub-test, there was a small difference (.29 of an SD) with men tending to score slightly higher than women. In contrast, a small difference (.23 of an SD) was found with women slightly outperforming men on the diagrammatic sub-test.

Age Group Differences

Total Score - Swift Technical Aptitude

Test	Under 40 N	Under 40 Mean	Under 40 SD	Over 40 N	Over 40 Mean	Over 40 SD	Pooled SD Difference
STA Total	2652	-.01	.56	1214	-.19	.53	.33

By Measure

Test	Test	Under 40 N	Under 40 Mean	Under 40 SD	Over 40 N	Over 40 Mean	Over 40 SD	Pooled SD Difference
Spatial*	STA sub-test	2652	.03	.76	1214	-.18	.73	.27
Mechanical	STA sub-test	2652	-.15	.68	1214	-.15	.69	.01
	Mechanical Reasoning Aptitude	1340	-.06	.90	634	-.10	.92	.04
Diagrammatic*	STA sub-test	2652	.10	.78	1214	-.23	.77	.42

No comparison on the Spatial Reasoning and Diagrammatic Reasoning single tests due to insufficient data from the Over 40 group (both N<500).

The tables above present the age group differences on the Swift Technical Aptitude Total Score, and separately the relevant tests that measure each of the three aptitude areas in Technical range – spatial, mechanical and diagrammatic.

Expressed in terms of raw theta (ability) scores, there was a small difference (.33 of a standard deviation) between the younger age group and the older age group on the Swift Technical Aptitude Total Score. Small differences were also found between the two groups on the Swift Technical Aptitude spatial and diagrammatic sub-tests (.27 and .42

of an SD respectively). In all of these cases, the younger age group overall scored slightly higher than the older age group. However, there was no notable difference between the two age groups on the Swift version or the full-length mechanical reasoning tests.

Ethnic Group Differences

Total Score - Swift Technical Aptitude

Test	White N	White Mean	White SD	Other Ethnicities N	Other Ethnicities Mean	Other Ethnicities SD	Pooled SD Difference
SCA Total	2564	.02	.53	1136	-.22	.55	.44

By Measure

Test	Test	White N	White Mean	White SD	Other Ethnicities N	Other Ethnicities Mean	Other Ethnicities SD	Pooled SD Difference
Spatial*	STA sub-test	2564	.06	.73	1136	-.24	.78	.39
Mechanical	STA sub-test	2564	-.06	.69	1136	-.32	.62	.39
	Mechanical Reasoning Aptitude	1392	.05	.92	608	-.41	.77	.52
Diagrammatic*	STA sub-test	2564	.05	.77	1136	-.11	.81	.21

**No comparison on the Spatial Reasoning and Diagrammatic Reasoning single tests due to insufficient data from the Other Ethnicities group (both N<500).*

The tables above present the ethnic group differences on the Swift Technical Aptitude Total Score, and separately the relevant tests that measure each of the three aptitude areas in Technical range – spatial, mechanical and diagrammatic.

Expressed in terms of raw theta (ability) scores, there was a small difference between the white group and other ethnicities on the Swift Technical Aptitude Total Score (.44 of a standard deviation). Small differences were found between the two groups on the Swift Technical Aptitude spatial and diagrammatic sub-tests (.39 and .21 of an SD respectively). For the Swift version and the full-length mechanical reasoning tests, the differences between the two groups were small to moderate (ranged from .39 to .52 of an SD). In all of these cases, the white group overall scored higher than other ethnicities.

Group Differences Summary

The data presented on the differences between the means for different groups reveal a number of group differences on the total scores. The differences here are broadly in line with the research literature and in some cases tend to be less pronounced than has been found in other studies. The differences on gender tend to be small (to almost non-existent). Small differences are seen on age with younger people, on average, achieving slightly higher overall scores. The largest differences are seen between white and other ethnicities with white groups on average performing higher.

The information presented here is from actual usage data of the Technical Aptitude Range and as a result the differences on some variables may reflect differences in the composition of the various groups. For example, age differences could be related to longer tenure in organizations and generational differences. Similarly, observed gender and ethnic differences could be a reflection of other biographical differences in the composition of these groups (e.g., level and type of education), rather than actual group differences.

Moreover, the performance differences reported are at the group level, rather than being reflective of specific individuals. In all cases, the average group-levels of performance represent largely overlapping performance distributions, with greater variation in performance within any group than between groups. Based on these average group-level data, it is inaccurate and inappropriate to make any predictions or decisions about any given individual's performance as a result of their membership of a particular ethnic group.

It is also important to bear in mind that each sample of individuals is different and group differences should not be generalized beyond these specifically-reported samples in an excessively broad manner. For example, the ethnic differences seen with cognitive tests are likely due, at least in part, to a difference in socioeconomic status and education across the specific groups of people sampled. While those group differences which do exist are interesting, it is worth noting that it is frequently difficult to isolate these variables as the sole determinant of the apparent difference.

As measures of cognitive ability, Technical Aptitude Range tests will occasionally reveal small to moderate differences between groups. To ensure that any group differences shown are meaningful, relevant and fair, it is important to make sure that the use of such tests can be justified. This is especially true when using a test in selection with a cut-off score. Justifying the use of any test involves making sure that the skills being assessed by the test are relevant and valid and that the level of any cut-off applied is demonstrably appropriate. The use of job analysis and, where possible, local validation studies is particularly important for demonstrating the link between a test and the job it is being used to select for.

In particular, the use of high cut-offs (e.g. above the 50^{th} percentile) may require additional justification and analysis to ensure that this does not lead to adverse impact against any group. A further precaution is to use a behavioral measure, e.g. Work Strengths or Match 6.5, alongside aptitude to create a weighted overall fit score which can be expected to mitigate against the potential for adverse impact.

It is one thing for an assessment to be designed to be fair and valid, and another for it to be used fairly. The clearer and more consistent the structure and process presented for aligning the Technical Aptitude Range to a job and agreeing consistent criteria for decision making based on the test, the less likely it is that the assessments will be unfairly applied by using different standards for candidates in different groups.

In general, the differences between age, gender and ethnic groups are small or moderate and we do not therefore advise that specific differences in profile interpretation should be warranted when considering test results from different groups defined according to these variables.

We do not, unless local legal frameworks permit or mandate such an approach, recommend using separate norms for age, gender or ethnic groups. For further information, please contact Saville Assessment directly.

10.0 Appendix 1: Internal Consistency Reliabilities for Sub-Tests in Swift Technical Aptitude

The following table shows the internal consistency reliability coefficients for the sub-tests in Swift Technical Aptitude.

Swift Technical Aptitude Internal Consistency Reliabilities (N=10682)

Sub-Test	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Spatial	57.94	19.36	1.25	6.25	.61
Mechanical	49.74	21.72	1.45	7.25	.47
Diagrammatic	52.46	24.93	1.21	6.05	.63

11.0 Appendix 2: Method for Calculating Criterion Related Validity of a Single Test from the Equivalent Swift Sub-Test in Swift Technical Aptitude

It is possible to calculate the criterion-related validity of each of the full-length, single tests based on validity evidence from the equivalent Swift sub-test. The variables used to derive the validity figures presented in this document are outlined below.

Test	Criterion Related Validity	Raw Criterion Related Validity of Equivalent STA Sub-Test (Epsom N=308)	Reliability of Equivalent STA Sub-Test*	Inter-Rater Reliability of STA Sub-Test's Equivalent Criterion (Epsom N=308)	Reliability of Single Tests**
Spatial Reasoning Aptitude	.18	.09***	.61	.33	.81
Mechanical Reasoning Aptitude	.52	.24	.47	.36	.80
Diagrammatic Reasoning Aptitude	.52	.19	.63	.18	.84
		(r_{xy})	(r_{xx})	(r_{yy})	(r_{zz})

* *Reliability figures for the equivalent Swift sub-tests are all internal consistency figures from the largest sample size available.*

** *Reliability figures for the single tests are all internal consistency figures from the largest sample size available.*

*** *Raw Spatial validity reported here uses the Spearman Brown Prophecy formula to estimate raw validity for 12 items based on the 8-item validity of .06 in Project Epsom = (1.5 * .06)/(1+((1.5-1) * .06))*