

Saville Assessment

WillisTowersWatson 



Analysis Aptitude Range Technical Summary



Contents

1.0	Analysis Aptitude Range Test Information.....	2
2.0	Norm Groups.....	3
3.0	Reports.....	3
4.0	Practice and Preparation.....	3
5.0	Development.....	3
6.0	Languages.....	3
7.0	Reliability.....	4
8.0	Validity.....	6
9.0	Fairness.....	9
10.0	Appendix 1: Development of Swift Executive Aptitude.....	16
11.0	Appendix 2: Development of Abstract Reasoning Aptitude.....	16
12.0	Appendix 3: Internal Consistency Reliabilities for Sub-Tests in Swift.....	18
13.0	Appendix 4: Method for Calculating Criterion Related Validity of a Single Test from the Equivalent Sub-Test in Swift Analysis/Executive Aptitude.....	19



1.0 Analysis Aptitude Range Test Information

Key information

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

- **Verbal Analysis** assesses the ability to understand, interpret and evaluate written information
- **Numerical Analysis** assesses the ability to understand, interpret and evaluate numerical data
- **Diagrammatic Analysis** assesses the ability to analyze diagrams, sequences and transformations
- **Abstract Reasoning** assesses the ability to understand sequences of patterns and relationships

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 Analysis Aptitude	Total	24	18
	Verbal Analysis	8	6
	Numerical Analysis	8	6
	Diagrammatic Analysis	8	6
Swift Executive Aptitude	Total	28	18
	Verbal Analysis	8	6
	Numerical Analysis	8	6
	Abstract Reasoning	12	6

Test	Total / Sub-Test	No. of Questions	Time Limit (mins)
Swift Analysis Verbal & Numerical	Total	32	24
	Verbal Analysis	16	12
	Numerical Analysis	16	12
Verbal Analysis Aptitude	Total	32	24
Numerical Analysis Aptitude	Total	32	24
Diagrammatic Analysis Aptitude	Total	32	24
Abstract Reasoning Aptitude	Total	32	16

Note: Supervised Access (SA) Analysis 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, regional and country specific norms are available for the tests in this range. Information on the 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 Analysis Aptitude Range consists of large banks of Verbal Analysis, Numerical Analysis, Diagrammatic Analysis and Abstract Reasoning items. Items are drawn from these banks to form the single Verbal, Numerical, Diagrammatic and Abstract tests and the corresponding sub-tests of Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical.

The development of Swift Executive Aptitude and Abstract Reasoning Aptitude can be found in Appendix 1 and Appendix 2 respectively.

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, 1982¹) 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 Analysis Aptitude Range tests.

For Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical, 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 above 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 Analysis Aptitude Internal Consistency Reliability (N=95337)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	60.67	19.62	.82	4.12	.83

¹Andrich, 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.

Swift Executive Aptitude Internal Consistency Reliability (N=22104)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	65.53	17.32	.85	4.24	.82

Swift Analysis Verbal & Numerical Internal Consistency Reliability (N=28647)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	67.33	18.84	.72	3.61	.87

Verbal Analysis Aptitude Internal Consistency Reliability (N=24740)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	74.49	16.48	.89	4.47	.80

Numerical Analysis Aptitude Internal Consistency Reliability (N=24518)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	63.35	17.43	.80	4.00	.84

Diagrammatic Analysis Aptitude Internal Consistency Reliability (N=8803)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	71.94	18.26	.75	3.74	.86

Abstract Reasoning Aptitude Internal Consistency Reliability (N=27856)

	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Total	69.74	17.22	.82	4.12	.83

8.0 Validity

Swift Analysis Aptitude

This summary document includes criterion-related validity information for the total score and three sub-tests in Swift Analysis Aptitude; based on the same 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 .69. 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 Verbal Analysis sub-test, with the least coming from the Diagrammatic Analysis sub-test.

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

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

	Correlation with Sum of Working with Words, Numbers and Systems (Rater)	
	Uncorrected r	Corrected r
Total	.29	.54
	Correlation with Working with Words (Rater)	
	Uncorrected r	Corrected r
Verbal	.27	.48
	Correlation with Working with Numbers (Rater)	
	Uncorrected r	Corrected r
Numerical	.20	.34
	Correlation with Working with Systems (Rater)	
	Uncorrected r	Corrected r
Diagrammatic	.10	.24

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 Words, Numbers and Details criteria (.31, .34 and .18 respectively). The criterion internal consistency of ratings (N=308) was .69.

Swift Executive Aptitude

The criterion-related validity information for the total score and three sub-tests in Swift Executive Aptitude is based on a sample of 214 high potential middle managers 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.

Validation Study of Middle Managers N=214

	Boss Rating			
	Overall (Average of 9 ratings)	Judgment (Average of 3 ratings)	Drive (Average of 3 ratings)	Influence (Average of 3 ratings)
Total	.32	.45	.26	.06
Total (raw)	.17	.24	.14	.03
Verbal	.08	.13	.08	-.01
Numerical	.15	.19	.15	.03
Abstract	.10	.16	.26	.04

Note: Any raw correlation higher than .16 is statistically significant at the $p < .01$ level (one-tailed) and any raw correlation higher than .11 is statistically significant at the $p < .05$ level (one-tailed). $N=214$.

As can be seen, in this study the strongest prediction of workplace performance came from the Numerical sub-test, with the Abstract and then Verbal sub-tests close behind. The total corrected validities have been adjusted by a criterion reliability estimate of .28. The raw validities have not been adjusted for any criterion unreliability or restriction of range. The total score correlates particularly well with the Overall, Judgment and Drive criteria, with a notable corrected correlation of .45 against the boss ratings of workplace Judgment.

Validity of Swift Executive Aptitude in Relation to Swift Analysis Aptitude

As the only difference between Swift Analysis Aptitude and Swift Executive Aptitude is the third sub-test (which is Diagrammatic or Abstract, respectively) and because the Abstract norm scores were originally calibrated from Diagrammatic norm scores (with a .72 correlation between the two test formats), the published Swift Analysis Aptitude validity (as given in the section above) provides a suitable additional source of evidence for the validity of Swift Executive Aptitude.

Swift Analysis Verbal & Numerical

The Swift Analysis Verbal & Numerical shares the same content bank as Swift Analysis Aptitude and has eight more items in each sub-test. Therefore, it is appropriate to assume that the validity figures for Swift Analysis Verbal & Numerical will be aligned to those presented & Numerical test, the Swift Analysis Aptitude validities should be considered a lower-bound estimate of the validities of Swift Analysis Verbal & Numerical.

Analysis Aptitude Range Single Tests

The Verbal Analysis, Numerical Analysis, Diagrammatic Analysis and Abstract Reasoning single tests are longer than the various Swift combined assessments and cover the same areas of aptitude in greater depth. It is appropriate to assume that the Swift Analysis/Executive Aptitude validities are a conservative and lower-bound estimate of the validity of the Analysis Aptitude single tests, which are likely to show incremental validity over the Swift assessments (see Appendix 4).

9.0 Fairness

Gender Group Differences

Total Score - Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical

Test	Male N	Male Mean	Male SD	Female N	Female Mean	Female SD	Pooled SD Difference
SAA Total	20482	-.10	.64	10931	-.27	.60	.27
SEA Total	6023	-.14	.65	3630	-.17	.60	.04
SAVN Total	5817	-.02	.76	4152	-.12	.77	.13

By Measure

	Test	Male N	Male Mean	Male SD	Female N	Female Mean	Female SD	Pooled SD Difference
Verbal	SAA sub-test	20482	-.13	.79	10931	-.23	.77	.14
	SEA sub-test	6023	-.09	.81	3630	-.11	.77	.02
	SAVN sub-test	5817	.01	.88	4152	.01	.87	.00
	Verbal Analysis Aptitude	7434	.28	.86	6201	.22	.86	.07
Numerical	SAA sub-test	20482	-.14	.79	10931	-.39	.74	.32
	SEA sub-test	6023	-.12	.79	3630	-.30	.75	.23
	SAVN sub-test	5817	-.06	.83	4152	-.26	.87	.23
	Numerical Analysis Aptitude	7557	-.11	.75	6094	-.38	.75	.36
Diagrammatic	SAA sub-test	20482	-.03	.83	10931	-.19	.78	.20
	Diagrammatic Analysis Aptitude	1429	.25	.90	617	.03	.82	.25
Abstract	SEA sub-test	6023	-.22	.86	3630	-.10	.82	-.14
	Abstract Reasoning Aptitude	13674	-.16	.91	6359	-.15	.91	-.01

The tables above present the gender group differences on the Total Scores for Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical, and separately the relevant tests that measure each of the four aptitude areas in Analysis range - verbal, numerical, diagrammatic and abstract.

Expressed in terms of raw theta (ability) scores, there was a small difference between males and females on the Swift Analysis Aptitude Total Score, but no notable difference on the Swift Executive Aptitude or Swift Analysis Verbal & Numerical Total Score.

In terms of individual measures, the verbal sub-tests in various Swift tests or the single Verbal Analysis test did not show any notable gender differences. There was also no notable difference between the two gender groups on the Swift version or the full-length Abstract Reasoning tests. However, there were small differences (ranged from .23 to .36 of a standard deviation) in the numerical sub-tests of Swift or the single Numerical Analysis test that males generally scored higher than females. The diagrammatic sub-test in Swift Analysis Aptitude and the single Diagrammatic Analysis test also show small gender differences (.20 and .25 of a standard deviation respectively) where males overall slightly outperformed females.

It is worthy of note that the small advantage for males on the numerical sub-test in Swift Executive Aptitude is compensated by the slightly higher average score for females on the abstract sub-test, meaning that effectively Swift Executive Aptitude shows no gender group difference on the Total Score.

Age Group Differences

Total Score - Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical

Test	Under 40 N	Under 40 Mean	Under 40 SD	Over 40 N	Over 40 Mean	Over 40 SD	Pooled SD Difference
SAA Total	22593	-.06	.63	6435	-.47	.54	.67
SEA Total	5206	.00	.61	3485	-.36	.59	.61
SAVN Total	5345	-.14	.69	1668	-.28	.68	.21

By Measure

	Test	Under 40 N	Under 40 Mean	Under 40 SD	Over 40 N	Over 40 Mean	Over 40 SD	Pooled SD Difference
Verbal	SAA sub-test	22593	-.08	.78	6435	-.44	.73	.47
	SEA sub-test	5206	.01	.78	3485	-.23	.80	.30
	SAVN sub-test	5345	-.01	.81	1668	-.18	.81	.20
	Verbal Analysis Aptitude	2936	.20	.90	3597	.18	.86	.02
Numerical	SAA sub-test	22593	-.15	.79	6435	-.49	.70	.44
	SEA sub-test	5206	-.08	.78	3485	-.33	.75	.33
	SAVN sub-test	5345	-.26	.77	1668	-.38	.75	.16
	Numerical Analysis Aptitude	3262	-.11	.80	3222	-.17	.74	.07
Diagrammatic	SAA sub-test	22593	.04	.81	6435	-.49	.70	.68
	Diagrammatic Analysis Aptitude	1329	.36	.86	507	-.27	.80	.74
Abstract	SEA sub-test	5206	.08	.80	3485	-.53	.76	.78
	Abstract Reasoning Aptitude	5602	-.14	.96	2256	-.57	.84	.47

The tables above present the age group differences on the Total Scores for Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical, and separately the relevant tests that measure each of the four aptitude areas in Analysis range - verbal, numerical, diagrammatic and abstract.

Expressed in terms of raw theta (ability) scores, moderate differences (.67 and .61 of a standard deviation) were found between the younger group and the older group on the Swift Analysis Aptitude and Swift Executive Aptitude Total Scores, but the age group difference on the Swift Analysis Verbal & Numerical Total Score was small (.21 of a standard deviation). In all these comparisons, the Under 40 group, overall, scored higher than the Over 40 group.

In terms of individual measures, the verbal sub-tests in various Swift tests show small age group differences (ranged from .20 to .47 of a standard deviation); the younger age group generally scored higher than their older counterparts, but there was no notable age difference on the single Verbal Analysis test. Similarly, the numerical sub-tests of Swift Analysis and Swift Executive Aptitude showed small age group differences (.44 and .33 of a standard deviation respectively) but there was no notable age difference on the numerical sub-test in Swift Analysis Verbal & Numerical or the single Numerical Analysis test. Both the Swift version and the full-length Diagrammatic Analysis tests showed moderate differences (.68 and .74 of a standard deviation respectively) where the younger age group, in general, outperformed the older age group. The Swift version and the full-length Abstract Reasoning tests also showed moderate age group differences with the younger group scoring higher overall (by .78 and .47 of a standard deviation respectively).

One possible contributor for the age group differences found in our Analysis Aptitude Range tests is that cognitive ability has been rising over time (the Flynn Effect). Much of this rise is attributed to the increase of fluid intelligence as the younger generations have a greater capacity to deal with abstract information (the digital generation). This can be seen in the age group comparisons presented above that the differences in the diagrammatic and abstract measures are larger than those found in verbal and numerical.

Another potential explanation for the age group differences is the difference in reaction times between younger and older generations. It is likely that younger generations are generally faster at answering questions, resulting in a higher completion rate on the test which increases the likelihood of achieving higher scores. However, our tests have been designed to have high completion rates to minimize the impact of reaction and completion times on scores.

Ethnic Group Differences

Total Score - Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical

Test	White N	White Mean	White SD	Other Ethnicities N	Other Ethnicities Mean	Other Ethnicities SD	Pooled SD Difference
SAA Total	11963	.01	.61	7075	-.30	.68	.49
SEA Total	4889	-.01	.57	3327	-.37	.67	.58
SAVN Total	4900	-.07	.66	3521	.02	.89	-.12

By Measure

	Test	White N	White Mean	White SD	Other Ethnicities N	Other Ethnicities Mean	Other Ethnicities SD	Pooled SD Difference
Verbal	SAA sub-test	11963	.07	.73	7075	-.36	.81	.56
	SEA sub-test	4889	.12	.71	3327	-.45	.79	.76
	SAVN sub-test	4900	.08	.78	3521	.02	.98	.07
	Verbal Analysis Aptitude	9470	.33	.83	742	-.34	.95	.81
Numerical	SAA sub-test	11963	-.10	.77	7075	-.36	.81	.33
	SEA sub-test	4889	-.10	.75	3327	-.32	.81	.29
	SAVN sub-test	4900	-.23	.75	3521	.02	.96	-.29
	Numerical Analysis Aptitude	9507	-.20	.75	860	-.44	.89	.32
Diagrammatic*	SAA sub-test	11963	.06	.80	7075	-.18	.85	.30
Abstract	SEA sub-test	4889	-.07	.81	3327	-.34	.89	.33
	Abstract Reasoning Aptitude	12253	-.06	.88	1867	-.67	.93	.69

*No comparison on Diagrammatic Analysis Aptitude due to insufficient data from the Other Ethnicities group (N<500).

The tables above present the ethnic group differences on the Total Scores for Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical, and separately the relevant tests that measure each of the four aptitude areas in Analysis range - verbal, numerical, diagrammatic and abstract.

Expressed in terms of raw theta (ability) scores, there were moderate differences between the white group and other ethnicities on the Swift Analysis Aptitude and Swift Executive Aptitude Total Scores (.49 and .58 of a standard deviation respectively). The white group generally scored higher on these tests than other ethnicities. However, the Swift Analysis Verbal & Numerical Total Score showed no notable difference between the ethnic groups.

In terms of individual measures, the Swift version or the full-length Verbal Analysis test show moderate to large ethnic group differences (ranged from .56 to .81 of a standard deviation) with the exception in Swift Analysis Verbal & Numerical sub-test having no notable difference. Small differences (ranged from .29 to .33 of a standard deviation) were found between the two groups on tests that measure Numerical Analysis. The white group outperformed other ethnicities on the numerical sub-test in Swift Analysis Aptitude and Swift Executive Aptitude, as well as the full-length Numerical Analysis test. However, other ethnicities overall scored higher than the white group on the numerical sub-test in Swift Analysis Verbal & Numerical. The difference between the two groups was small (.30 of a standard deviation) on the diagrammatic sub-test in Swift Analysis Aptitude, and the tests that measure Abstract Reasoning showed small to moderate ethnic group differences (ranged from .33 to .69 of a standard deviation). In all of these cases, the white group generally scored higher than other ethnicities.

These results are consistent with ethnic group difference findings for cognitive ability tests, with the differences for Verbal Analysis generally being the largest.

The ethnic group differences found in Swift Analysis Verbal & Numerical appeared to be different from other tests that also measure the verbal and numerical constructs. The Other Ethnicities group overall outperformed their White counterparts on the SAVN Total Score and the numerical sub-test, and there was no difference in the verbal sub-test. These different patterns can be explained by the composition of the Other Ethnicities group in the data for this particular test that there was a high proportion of Asian ethnicities and in particular Chinese who the wider research literature indicates, on average, are slightly higher performers.

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 Analysis 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, education and language proficiency 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, Analysis 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 50th 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 Analysis 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: Development of Swift Executive Aptitude

Development of the Swift Executive Aptitude assessment began in 2009. It is based upon the Swift Analysis Aptitude assessment, however the third sub-test, Diagrammatic Analysis, has been replaced with Abstract Reasoning which is another measure of systematic logical reasoning. The verbal and numerical analysis items for Swift Executive Aptitude are taken directly from the large banks that power Swift Analysis Aptitude and are therefore directly comparable across the tests.

Based on analysis of data from the Diagrammatic tests, as well as client feedback, it was suggested that an abstract format of logical reasoning might be more face-valid and appropriate for some roles (e.g., senior and executive roles). The development of the abstract series format had the aim of producing a straightforward, language-free and globally applicable alternative to diagrammatic format tests, while still measuring the same performance criterion of 'Working with Systems'. The Abstract Series format also has the advantage that it requires less explanatory text, so is particularly suitable for translation and global use. Please refer to Appendix 2 for more information about the development of the Abstract Reasoning test.

The Swift Executive Aptitude assessment is formed of three sub-tests and takes a total time of 18 minutes to complete (six minutes per sub-test). The updated 2011 version features enhanced scoring based on the principles of Item Response Theory (IRT), and large banks of randomized item content powered by an increased number of norm groups compared to that of the original 2009 version. It also benefits from pace information, which is an enhanced method of measuring a test taker's completion rate. This score is displayed in the Aptitude & Pace Report which is available for all randomized tests.

11.0 Appendix 2: Development of Abstract Reasoning Aptitude

Development of an Abstract test format began in 2009. Based on analysis of data from the Diagrammatic tests, as well as client feedback, a new format for measuring systematic logical reasoning was targeted. Client feedback had suggested that an abstract format of logical reasoning might be more face-valid and appropriate in some roles (e.g., senior and executive roles). Saville Assessment began development of an Abstract Series format with the aim of producing a straightforward, language-free and globally applicable alternative to diagrammatic format tests, while still measuring the same performance criterion of 'Working with Systems'. The Abstract Series format also has the advantage that it requires less explanatory text and so is particularly suitable for translation and global use.

Items were written by members of the Research & Development team with a view to capturing established constructs of abstract reasoning which underpin a range of existing abstract format tests. As with all of the Saville Assessment aptitude tests, different item types were developed in order to capture multiple specific abilities within the overall construct of abstract reasoning. In order to be consistent with the Diagrammatic tests, which are designed to measure similar performance constructs, the item types used in creating the abstract items are similar to some of the item types used in the diagrammatic content: 'Identifying Rules', 'Comprehending Processes' and 'Understanding Logical Sequences'. The difference between item types for the abstract content is dependent on where the question mark (?) sits in the abstract series. If the '?' is at the beginning of the series it is 'Identifying Rules'; if the

'?' is in the middle of the series it is 'Comprehending Processes' and if the '?' sits at the end of the series it is 'Understanding Logical Sequences'. Testlets were constructed and arranged to sample these multiple item types and to get progressively harder, both within a testlet and across the test as a whole.

Trialing

The initial trialing phase consisted of a group of paid volunteers completing a range of new and existing aptitude content including 64 abstract items. The time allowed for each testlet of four questions was three minutes. This was to allow for time vectoring, whereby trial candidates are given generous time limits in order to analyze the effect of different completion times and to determine the best time limit for a test. Following the trialing, item analysis was conducted to select the best items for the first version of Abstract Reasoning. A correlation of .72 was found between the Abstract Series and Diagrammatic Reasoning formats, confirming the high degree of similarity between these two formats, and providing reassurance that they are both valid assessments of systematic, logical reasoning. On the strength of this correlation, norms from the Diagrammatic test were calibrated across to the Abstract version so that they could be used with the new test, which was initially marketed as Global Abstract Series. This test was scored using the principles of Item Response Theory (IRT 3-Parameter model) and featured 36 items presented over 18 minutes.

In 2011-2012, international trialing took place using paid volunteers from a range of cultures and backgrounds. They completed a large number of aptitude tests including technical and comprehension assessments which were allocated to them based on their stated job roles. On the basis of this trialing, an updated version of the Abstract format was created, with an increased item bank size. This version, released in 2012, is known as Abstract Reasoning Aptitude and features 32 items drawn from a large bank presented in 16 minutes.

Each test comprises eight timed testlets. Each testlet comprises four items and has a time limit of two minutes. The time limit on each testlet helps the candidate to pace themselves throughout the test and helps to prevent a candidate being heavily penalized if they spend a long time on one question.

Testlets are drawn from a large testlet bank. All testlets are arranged into different levels of difficulty in the bank (for example, 'Easy', 'Medium' and 'Difficult'). For the system to build a test for a specific candidate, there are rules in place to ensure that the test remains highly reliable and also to prevent over-exposure to questions of similar difficulty. These rules are to balance which items are presented from the bank. This is accomplished by the system requiring a test to be built with testlets of specific levels of difficulty in each testlet position throughout the test. The first testlet position is graded as 'Easy' and the final testlet position is graded as 'Difficult' for example. Therefore, in the first position only testlets graded as 'Easy' can be drawn from the bank and for the final position only testlets graded as 'Difficult' can be drawn from bank. Consequently, each candidate is presented with a randomized test but with a closely controlled and very similar level of difficulty.

Final equivalence is provided for all candidates using IRT to equate scores to ensure that all candidates are measured on a consistent and comparable scale.

Abstract Reasoning Aptitude features enhanced IRT scoring and a larger bank of randomized item content powered by an increased number of norm groups. It also benefits from pace information, which is an enhanced method of measuring a test taker's completion rate. The Pace score is based on the candidate's response time for the questions they completed compared to the average response time for the same questions. It can be defined as how quickly a candidate has responded compared to the average for the same questions. This score is displayed in the Aptitude & Pace Report available for all randomized tests.

A supervised access online parallel form of Abstract Reasoning Aptitude was developed in 2014. This provides the reassurance of having a supervised follow-up test available if so required. As is consistent with the other Supervised Access tests in the Saville Assessment Aptitude test portfolio, this test is scored using the principles of Classical Test Theory and features fixed content. This will permit the creation of a parallel hardcopy supervised version, which can be hand scored, in due course. Following trialing on a group of paid volunteers and live client usage in 2013, the test was made commercially available in 2014 with an initial standardization norm. Further norms will be released in due course.

For further information about Saville Assessment's development processes and for an in-depth explanation of the principles of IRT, please refer to the Analysis Aptitude Range handbook.

12.0 Appendix 3: Internal Consistency Reliabilities for Sub-Tests in Swift

The following tables show the internal consistency reliability coefficients for the sub-tests in Swift Analysis Aptitude, Swift Executive Aptitude and Swift Analysis Verbal & Numerical. The sub-test scores provide additional test-taking information and should not be used in isolation for decision making. Therefore, we are not seeking the sub-tests to have reliability estimates higher than .70.

Swift Analysis Aptitude Internal Consistency Reliabilities (N=95337)

Sub-Test	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Verbal	63.62	24.16	1.23	6.16	.62
Numerical	53.60	25.42	1.13	5.66	.68
Diagrammatic	64.80	24.20	1.18	5.92	.65

Swift Executive Aptitude Internal Consistency Reliabilities (N=22104)

Sub-Test	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Verbal	66.98	23.53	1.25	6.24	.61
Numerical	56.28	24.88	1.15	5.74	.67
Abstract	70.74	19.74	1.13	5.66	.68

Swift Analysis Verbal & Numerical Internal Consistency Reliabilities (N=28647)

Sub-Test	Mean % Correct	SD (%)	SEm Sten	SEm 'T'	r
Verbal	69.39	20.89	1.00	5.00	.75
Numerical	65.27	21.15	0.89	4.47	.80

13.0 Appendix 4: Method for Calculating Criterion Related Validity of a Single Test from the Equivalent Sub-Test in Swift Analysis/Executive 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.

Verbal Analysis, Numerical Analysis and Diagrammatic Analysis - based on Swift Analysis Aptitude (SAA)

Test	Criterion Related Validity	Raw Criterion Related Validity of Equivalent SAA Sub-Test (Epsom N=308)	Reliability of Equivalent SAA Sub-Test*	Inter-Rater Reliability of SAA Sub-Test's Equivalent Criterion (Epsom N=308)	Reliability of Single Tests**
Verbal Analysis Aptitude	.55	.27	.62	.31	.80
Numerical Analysis Aptitude	.38	.20	.68	.34	.84
Diagrammatic Analysis Aptitude	.27	.10	.65	.18	.86
		(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.

Abstract Reasoning – based on Swift Executive Aptitude (SEA)

Test	Criterion Related Validity	Raw Criterion Related Validity of Equivalent SEA Sub-Test (Middle Managers N=214)	Reliability of Equivalent SEA Sub-Test*	Inter-Rater Reliability of SEA Sub-Test's Equivalent Criterion (Middle Managers N=214)	Reliability of Single Test**
Abstract Reasoning Aptitude	.21	.10	.68	.28	.83
		(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.

Saville Assessment

CI Tower
1st Floor
St George's Square
New Malden
KT3 4HG
United Kingdom

Tel +44(0)20 8619 9000

info@savilleassessment.com