12.0 Validity

In order to allow us to make better decisions about people based on assessments they have completed, it is essential that the assessments can provide evidence of their validity.

This chapter is divided into three parts. First is an introduction to, and orientation in, the concept of validity. This introduction helps explain why the Analysis Aptitude assessments are different from traditional aptitude assessments in certain respects. While validity itself can be thought of as a unitary concept, this chapter will cover two aspects of validity with regard to the Analysis Aptitude Range: criterion-related validity and construct validity.

For a summary of the criterion-related and construct validity of the Analysis Aptitude Range, please refer to section 12.7 at the end of this chapter.

12.1 What is Validity?

How do we know that a test or an assessment works and actually does what it claims to do? How do we know if the inferences and decisions made using one assessment are any better than another? How can we know which assessment tools are most accurate and will maximize the amount of benefit derived from the information provided? Which tool is the best investment for an organization?

These and other related questions are fundamental to the development and continuous improvement of assessment in the workplace. These questions directly relate to the validity of an assessment. All Saville Consulting aptitude assessments are designed to maximize their validity in forecasting effectiveness at work, in terms of both overall performance and in relation to specific workplace competencies.

The validity of an assessment is its most important property. For an assessment developed for use in the workplace, validity is essentially concerned with how well the assessment actually relates to, or predicts, the relevant aspects of work performance. Even if an assessment can accurately measure an obscure psychological theory, this is of little importance if the assessment is unrelated to performance at work. Assessments, such as aptitude tests, used to make selection and performance decisions in the workplace must forecast the important aspects of workplace performance. The interpretation of scores on an assessment, as well as the resulting inferences and decisions, should be valid.

In the words of the *Standards for Educational and Psychological Testing* (1999), of the American Psychological Association:

"Validity is a unitary concept. It is the degree to which all the accumulated evidence supports the intended interpretation of the test scores for the proposed purpose."

This chapter considers two different forms of accumulated evidence: criterion-related and construct validity.

Criterion-related validity is often seen as one of the most important properties of an assessment. For our application, it is the relationship between a score on an assessment and the relevant measure of performance at work. The criterion-related validity shows an assessment's appropriateness for the intended application. It is typically assessed by having independent raters judge the performance of the test-taker at work and then evaluating the strength of the relationship between these judgements and scores on the test.

Where independent performance ratings are provided at the same time as the test is completed, concurrent validity is being assessed. Measurement of the relationship between an assessment score and a later measure of job performance assesses predictive validity.

Construct validity is the extent to which an assessment measures a hypothetical construct or area of human performance. The scores from an assessment with good construct validity would be expected to behave as if the underlying construct was directly being measured. In essence, an assessment needs to provide a good reflection of the construct it is measuring. Where constructs are well-defined and used, such as in an aptitude assessment, a common and useful method for measuring construct validity is by comparison with other assessments which aim to measure the same or similar constructs.

Correlations between different assessments are presented to support the construct validity of the Analysis Aptitude assessments.

Construct validity focuses on convergent and discriminant evidence, looking for instances where relationships which are expected to occur are present (convergent evidence), and instances where relationships which are not expected to occur are absent (discriminant evidence). By pre-hypothesizing where we would expect to see relationships between two or more scales and where we would not, we can build up an idea about the construct validity of the assessment in question.

For example, if an established personality questionnaire has a scale measuring 'reasoning', we would expect a person's score on this scale to correlate with their scores on Saville Consulting's aptitude tests. However, it is perhaps less likely that there will be a strong relationship between a person's level of optimism and their score on the Saville Consulting aptitude tests. In fact the presence of such a relationship could point towards construct-irrelevant variance being present in the aptitude test.

Amassing construct validity evidence is therefore a continuing scientific pursuit to build up a body of evidence about how an assessment instrument works in practice, rather than simply something that a test possesses or does not possess (Landy, 1985).

It is worth mentioning two other types of validity which are relevant to assessments: content validity, which looks at how representative the questions in an assessment are of real workplace tasks; and face validity, a measure of how job-relevant an assessment is perceived, at face value, to be. Although these aspects are commonly measured through less quantitative methods than construct validity, they are, nonetheless, very important, since they provide an indication of an assessment's relevance to the content domain in question.

12.2 Validity and Inferences from Assessment Scores

Validity has at its heart the notion of the validation of inferences made from assessment scores. Many inferences can be explicitly stated as hypotheses and tested empirically. For example, it may be believed that individuals who score highly on a particular aptitude assessment scale are likely to be more analytical than people who score lower on the scale. Testing such inferences can provide information that a scale is meaningful and that drawing inferences about a person's analytical ability based on such a scale is appropriate. Assessments which are demonstrated to be valid can be used to make specific inferences about people's behaviors and likely performance.

Each assessment within the Analysis Aptitude Range is designed to predict a different area of workplace performance. Verbal Analysis is designed to measure potential for Working with Words, Numerical Analysis is designed to measure potential for Working with Numbers and Diagrammatic Analysis is designed to measure potential for Working with Systems. These three criterion areas form part of the Saville Consulting Wave Performance Framework.

In all instances, criterion-related validity is crucial. We believe that assessment users benefit from having clear links between the predictors (the assessment) and the criteria (workplace performance). The validity of an assessment helps ensure that it is used appropriately. The justification for using an assessment is derived partly from the extent to which it can measure the required criteria of workplace performance.

Such evidence provides us with confidence that the item is meaningful and the inference or interpretation we then draw using the item score is appropriate. General inferences such as these can be explicitly stated as hypotheses and tested empirically.

If a test or assessment score is shown to be a valid measure of a technical skill, such as mechanical reasoning, and if we learn from a job analysis study that a skill like mechanical reasoning is important to successful job performance, then we may want to use the test or assessment score to identify which people are more adept at mechanical reasoning than others.

12.3 Validity and Project Epsom

Beginning in 2007, Saville Consulting began a large-scale co-validation study (Project Epsom), in order to establish and compare the validity of a number of commonly-used assessments. A matched sample (N=308) from a diverse range of job sectors completed an assortment of personality measures. Independent ratings of their work performance were also obtained. This included the fixed-form online unsupervised versions of the Swift Technical Aptitude, Swift Comprehension Aptitude and Swift Analysis Aptitude tests. Project Epsom therefore contributed to the evaluation of the construct validity of the Swift Analysis Aptitude assessment, by comparing correlations between its scales and those of a range of other measures.

Some of the personality measures not developed by Saville Consulting which were included in Project Epsom are:

OPQ32i, NEO-PI-R, the Hogan Personality Inventory (HPI), 16PF-5, the Myers-Briggs Type Indicator (MBTI), Firo-B, the Hogan Development Survey (HDS), Axiom Software's DISCUS, the Thomas International Personal Profile Analysis (DISC) and the 15FQ+. (The Thomas International Personal Profile Analysis and 15FQ+ questionnaire were completed by a smaller number of participants as they were presented as additional questionnaires which became available part-way through Project Epsom).

Two personality measures developed by Saville Consulting were also included in Project Epsom: Wave Professional Styles and Wave Focus Styles.

The rest of this chapter details various forms of validity evidence arising from research on the assessments in the Analysis Aptitude Range.

12.4 Criterion-Related Validity

For a summary of the criterion-related validity of the Analysis Aptitude Range, please refer to section 12.7 at the end of this chapter.

Concurrent Criterion-Related Validity of Swift Analysis Aptitude from Project Epsom

Using the data on 308 participants from Project Epsom, the total scores of each of the three sub-tests in Swift Analysis Aptitude were combined, as were the independent third-party ratings of effectiveness in the matched areas of workplace performance. These data are shown in the table below.

Table 12.1 Concurrent Validity of the Swift Analysis Aptitude Overall Total Score Against the Sum of the Matched Work Performance Criteria (N=308)

	Swift Analysis Aptitude (Verbal, Numerical and Diagrammatic)
Validity Rater (Adj. for Inter-rater r) (N=308)	**.54**
Validity (raw) Rater (N=308)	.29
Matched Criterion Rating	Sum of Ratings for Working with Words, Numbers and Systems
Predictor Mean	11.43
Predictor SD	3.78
Criterion Mean	15.69
Criterion SD	2.83

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 (N=263) was .29 and the criterion internal consistency of ratings (N=308) was .69.

The table opposite provides a summary of the concurrent validity of the Swift Analysis Aptitude assessments against key aptitude-related competencies, as evidenced from Project Epsom. The raw total validity estimate is .29. Once 'corrected' for the inherent limitations of errors and inconsistencies in the ratings of people's effectiveness at work, the total validity estimate is .54.

The internal consistency of the summed criterion used is .69, which 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 is always going to be some degree of difference between multiple raters' judgements of effectiveness on the criteria of interest, which can force the validity coefficient down.

Here, inter-rater reliability at .29 is low, suggesting there is not a high level of agreement between the two raters on how effective the respondent is. Whether we use the raw or adjusted validity estimate, we can conclude that the Swift Analysis Aptitude demonstrate a respectable level of criterion-related validity, as evidenced from Project Epsom.

The Criterion-Related Validity of the Analysis Aptitude -R Single Tests

The Verbal, Numerical and Diagrammatic Analysis Aptitude -R single tests are longer than the Swift Analysis Aptitude combined assessment and cover the same areas of aptitude in greater depth. It is appropriate to assume that the Swift Analysis Aptitude validities are a conservative and lower-bound estimate of the validity of the Analysis Aptitude -R single tests, which are likely to show incremental validity over the Swift assessment.

Criterion-Related Validity of -R Single Analysis Aptitude Tests

Because the -R single tests and equivalent -R Swift version share the same bank(s) of test content, it is possible to provide summary validity figures for the Analysis -R single tests which take into account their greater length and reliabilities.

The figures below provide a summary of the criterion-related validities of each of the -R single tests which corrects the Swift figures for the appropriate full-length test reliabilities.

Table 12.2 Criterion-Related Validity of Analysis Aptitude -R Single Tests

Test	Criterion-Related Validity*
Verbal Analysis Aptitude-R	.59
Numerical Analysis Aptitude-R	.40
Diagrammatic Analysis Aptitude-R	.28

**Corrected for full length predictor reliability and criterion reliability*

For details about the method for calculating criterion-related validities of the -R single tests, please see Appendix C.

Concurrent Criterion-Related Validity of Swift Analysis Aptitude Sub-Tests in Project Epsom

Each of the sub-tests in the Swift Analysis Aptitude assessment were also correlated individually with their respective matched workplace competency. These data from Project Epsom are shown in the table below.

Table 12.3 Concurrent Validity of Swift Analysis Aptitude Sub-Test Total Scores Against the Sum of the Matched Work Performance Criteria (N=308)

	Correlation with Working with Words (Rater) r (Uncorrected)	Correlation with Working with Words (Rater) r (Corrected)
Verbal Total	**.27**	.48
	Correlation with Working with Numbers (Rater) r (Uncorrected)	**Correlation with Working with Numbers (Rater) r (Corrected)**
Numerical Total	**.20**	.34
	Correlation with Working with Systems (Rater) r (Uncorrected)	**Correlation with Working with Systems (Rater) r (Corrected)**
Diagrammatic Total	**.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 Systems criteria (.31, .34 and .18 respectively). The criterion internal consistency of ratings (N=308) was .69.

As shown in the table above, the total scores for each of the three sub-tests in the Swift Analysis Aptitude assessments were correlated with their respective matched criterion scales in the Saville Consulting Wave Framework of Performance and Potential. The criterion scores for each individual were estimated by a third-party rater of their workplace performance. The corrected correlation figure for the Verbal sub-test (.48) is a good deal higher than that for the Diagrammatic sub-test (.24).

Swift Analysis Aptitude Correlations with Performance Ratings of Accountants

178 trainee chartered accountants for a major professional organization completed Swift Analysis Aptitude (Invited Access version) and this was correlated with their scores on a professional accountancy exam. The correlations between the total overall Swift Analysis Aptitude score and the scores on each of the different units of the accountancy exam are shown below.

Table 12.4 Correlations between Swift Analysis Aptitude and Accountancy Exam Scores (N=178)

Accountancy Exam Module	Correlation with Swift Analysis Aptitude Overall Total Score (Uncorrected)	Correlation with Swift Analysis Aptitude Overall Total Score (Corrected)
Taxation	.21	.27
Advanced Finance	.37	.48
Assurance and Business Systems	.17	.22
Financial Reporting	.22	.28
Business Law	.23	.30
Business Management	.18	.23
Finance	.32	.41
Principles of Auditing and Reporting	.16	.21
Financial Accounting	.32	.41

Note: Any raw correlation higher than .15 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .13 is statistically significant at the p<.05 level (one-tailed). N=178. The corrected figures are adjusted according to an attenuation estimate of .60.

As can be seen in the table above, the Swift Analysis Aptitude overall total score correlated reasonably closely with the accountancy exam scores (.16 to .37 uncorrected).

Swift Analysis Aptitude Correlations with Selection Centre Outcomes of Medical Trainee Applicants

Swift Analysis Aptitude (Invited Access) data were collected from a sample of 215 candidates who were applying for training into General Practice in 2009.

Of the 215 who completed the SAA there were 13 who did not consent to their pilot data being matched to their live selection data and so they were removed from the validation analysis reducing the sample to N=202.

Correlations were carried out between SAA scores and selection center exercise scores. It should be noted that 14 candidates from this sample did not get through to the later stage of the selection process, reducing the 'Stage 3' data to N=188.

Table 12.5 Correlations between Swift Analysis Aptitude (IA) and Selection Centre Scores (N=188)

	Group Exercise	Written Exercise	Simulation Exercise	Empathy & Sensitivity Score	Communication Skills Score	Problem Solving Score	Professional Integrity Score	Coping with Pressure Score	Clinical Expertise Score	Stage 3 Total Score
SAA Overall Total Corrected	**.36**	**.30**	**.41**	**.39**	**.37**	**.34**	**.46**	**.46**	**.50**	**.50**
SAA Overall Total Uncorrected	**.28**	**.23**	**.32**	**.30**	**.29**	**.26**	**.36**	**.36**	**.39**	**.39**
SAA Verbal Total	.19	.16	.37	.29	.22	.16	.32	.32	.36	.35
SAA Numerical Total	.23	.23	.28	.28	.26	.24	.31	.32	.36	.35
SAA Diagrammatic Total	.16	.12	.10	.13	.12	.20	.14	.15	.24	.18

Note: Any raw correlation higher than .15 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .13 is statistically significant at the p<.05 level (one-tailed). N=188. The corrected figures are adjusted according to an attenuation estimate of .60.

Correlation analyses indicated that there was a positive correlation between the Swift Analysis Aptitude and all the exercises; group, written and simulation at .28, .23 and .32 respectively. The overall Stage 3 Total Score and the Clinical Expertise Score correlated .39 with the SAA Overall Total Score.

Swift Analysis Aptitude Correlations with Performance Ratings in an IT Company

128 IT services employees completed the Swift Analysis Aptitude assessments. In addition, their line manager was asked to rate their work performance on the Wave Performance 360. The correlations between these scores are shown below.

Table 12.6 Correlations between Swift Analysis Aptitude and Ability Dimension Scores from Line Manager Ratings of Performance (N=128)

	Working with Words (Uncorrected)	Working with Numbers (Uncorrected)	Working with Details (Uncorrected)	Working with Systems (Uncorrected)	Working with Designs (Uncorrected)	Working with Equipment (Uncorrected)
Total	.17	.19	.18	.26	.11	-.03
Verbal	.21	.07	.07	.08	.17	.21
Numerical	.12	.17	.14	.17	.18	-.10
Diagrammatic	.03	.16	.17	.28	.11	-.16

Note: Any raw correlation higher than .18 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .15 is statistically significant at the p<.05 level (one-tailed). N=128. These figures have not been corrected for any criterion unreliability or restriction of range in the sample.

The table above shows that the Swift Analysis Aptitude sub-scores predicted reasonably well the line manager ratings on the three structurally-aligned ability dimensions 'Working with Words', 'Working with Numbers' and 'Working with Systems'.

12.5 Construct Validity

For a summary of the Construct Validity of the Analysis Aptitude Range, please refer to section 12.7 at the end of this chapter.

Professional Aptitudes Correlations with Educational Criteria

Scores from the Professional Aptitudes standardization sample were correlated with UK GCSE point scores for English, Mathematics and Science for 227 individuals. (Grades of A* were equated to 8 points; A = 7 points; B = 6 points; C = 5 points; D = 4 points; E = 3 points; F = 2 points and G=1 point). An overall score for these three GCSE assessments was also created by combining the three separate point scores. These correlations are shown below.

Table 12.7 Correlations between Professional Aptitudes Scores and UK GCSE English, Mathematics, Science and Combined Criteria (N=227)

	GCSE Combined Score	GCSE English Score	GCSE Mathematics Score	GCSE Science Score
Overall Total	.53	.38	.52	.42
Verbal Total	.49	.47	.39	.37
Numerical Total	.42	.28	.46	.31
Diagrammatic Total	.38	.16	.42	.35

Note: Any raw correlation higher than .14 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .11 is statistically significant at the p<.05 level (one-tailed). N=227.

The criterion-related validity of the Professional Aptitudes assessments is supported by strong correlations with GCSE point scores. The Professional Aptitudes overall total score correlated the strongest with the combined GCSE points score (.53). The Verbal and Numerical Aptitude scores showed strong correlations with their related GCSE subjects, English and Mathematics (.47 and .46 respectively). These figures have not been corrected for any criterion unreliability or restriction of range in the sample.

Construct Validity of Work Aptitudes from the Standardization Sample

Scores from the Work Aptitudes standardization sample were correlated with GCSE point scores for English, Mathematics and Science, for 273 individuals. (Grades of A* were equated to 8 points; A = 7 points; B = 6 points; C = 5 points; D = 4 points; E = 3 points; F = 2 points and G=1 point). An overall score for these three GCSE assessments was also created by combining the three separate point scores. These correlations are shown in the table below.

Table 12.8 Correlations between Work Aptitudes Scores and UK GCSE English, Mathematics, Science and Combined Criteria (N=273)

	GCSE Combined Score	GCSE English Score	GCSE Mathematics Score	GCSE Science Score
Overall Total	.62	.38	.60	.56
Verbal Total	.56	.43	.45	.52
Numerical Total	.50	.23	.55	.47
Diagrammatic Total	.50	.32	.51	.44

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=273.

The criterion-related validity of the Work Aptitudes assessments is supported by strong correlations with GCSE point scores. The Work Aptitudes overall total score correlated the strongest with the combined GCSE points score (.62).

In the prediction of the GCSE combined score, the subset of Work Aptitudes scores which form the Swift Analysis Aptitude hard-copy and supervised access version achieved an overall total validity of .53, while the Verbal, Numerical and Diagrammatic scores were .41, .42 and .42 respectively. This suggests that about 80% of the overall Work Aptitudes validity can be achieved with just 30% of the items. These figures have not been corrected for any criterion unreliability or restriction of range in the sample.

Comparing Pace against Speed, Accuracy and Caution in the Analysis Aptitude Range

As noted in the 'Test-Taking Information and Item Types' chapter, the randomized (-R) tests in the Analysis Aptitude Range now feature the new online Pace score which makes use of the latest technology to deliver a completion rate measure that is suitable for randomized online tests. The following section provides details to help users understand the relationship between Pace and the other scores within the Analysis Aptitude Range.

Table 12.9 Analysis Aptitude Range Pace, Speed, Accuracy, Caution and Theta Descriptive Statistics

	Pace			Original Test-Taking Styles			Theta			
	Verbal	Numerical	Diagrammatic	Speed	Accuracy	Caution	Overall	Verbal	Numerical	Diagrammatic
N	27,852	27,860	27,861	37,482			37,459			
Mean	-.03	-.04	-.20	.96	.74	-.22	.18	-.04	.29	.29
Std. Deviation	.23	.21	.28	.07	.15	.15	.56	.72	.76	.74
Minimum	-3.56	-3.42	-4.50	.04	.00	-.92	-2.15	-2.21	-2.22	-2.36
Maximum	.95	.97	.96	1.00	1.00	.83	1.50	1.41	1.35	1.75

Table 12.10 Intercorrelations of Swift Analysis Aptitude -R Pace, Speed, Accuracy Caution, and Theta Sores (N=27830-37482)

		Pace			Original Test-Taking Styles				Theta			
		Verbal	Numerical	Diagrammatic	Speed	Speed Corrected*	Accuracy	Caution	Overall	Verbal	Numerical	Diagrammatic
Pace	Verbal		.31	.26	.43	.89	.19	.00	.29	.33	.20	.14
	Numerical	.31		.38	.52	.93	.35	.12	.46	.32	.45	.29
	Diagrammatic	.26	.38		.41	.88	.28	.09	.35	.23	.23	.35
Test-Taking Style	Speed	.43	.52	.41		N/A	.28	-.17	.50	.39	.43	.32
	Speed Corrected*	.89	.93	.88	N/A		.77	-.59	.92	.87	.89	.82
	Accuracy	.19	.35	.28	.28	.77		.90	.94	.68	.75	.70
	Caution	.00	.12	.09	-.17	-.59	.90		.73	.52	.57	.57
Theta	Overall	.29	.46	.35	.50	.92	.94	.73		.73	.80	.75
	Verbal	.33	.32	.23	.39	.87	.68	.52	.73		.39	.28
	Numerical	.20	.45	.23	.43	.89	.75	.57	.80	.39		.42
	Diagrammatic	.14	.29	.35	.32	.82	.70	.57	.75	.28	.42	

**Validities corrected for restriction of range in the Speed measure (estimated SD of Speed Corrected is four times the Speed Restricted SD)*

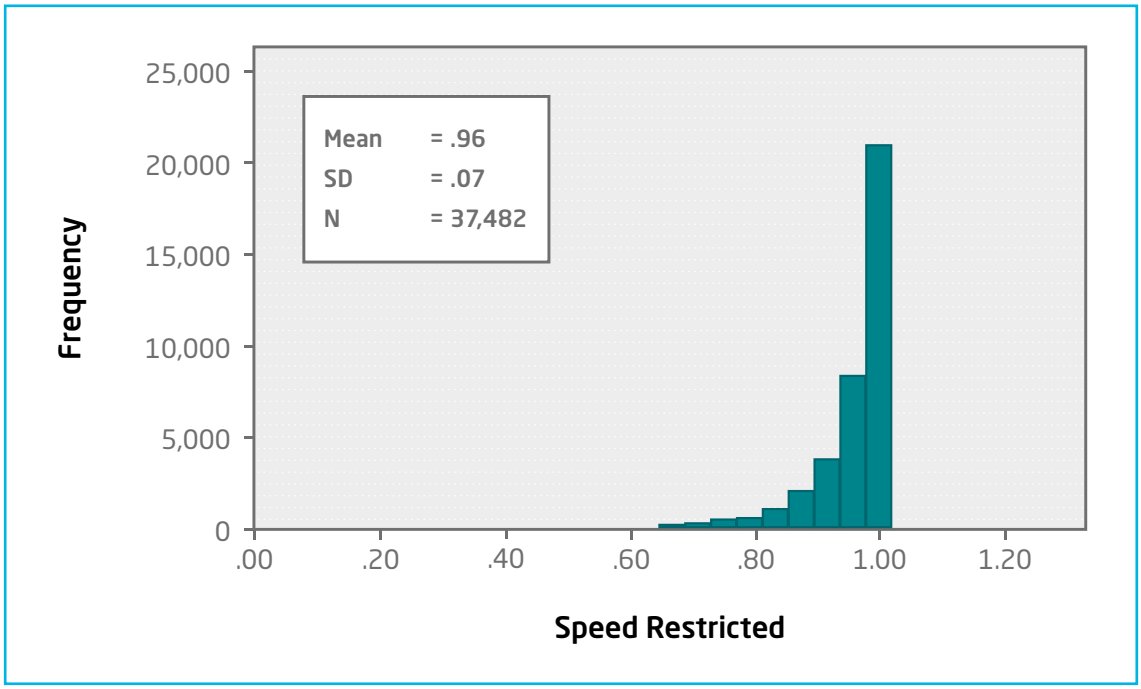
As can be seen from the intercorrelations above, in a large sample of live Swift Analysis Aptitude usage, the Pace scores correlate in the range of .26 to .38 amongst themselves. This suggests a relative degree of independence of measurement across the three areas. It is therefore possible for candidates to show differential patterns of Pace across the three sub-tests within Swift Analysis Aptitude, which may be particularly useful for diagnostic purposes during the feedback of scores.

Pace scores also correlate with the total (theta) score on Swift Analysis Aptitude-R in the range of .29 to .46 and in the range of .14 to .45 at the sub-test level. This suggests that while completion Pace has some positive relationship with aptitude (in general candidates who work at a faster pace tend to score better on the test), there is a relative degree of independence in these scales (therefore, some candidates are likely to score well while working at a lower pace, and vice versa).

Of particular note is that the Pace scores also correlate in the range of .41 to .52 with the Speed score, suggesting there is a degree of overlap in what these two completion rate measures are measuring. This reassures that meaningful interpretations are possible for a candidate whose Pace data are available at the unsupervised stage of assessment and Speed data for a subsequent follow-up in a Hard-Copy or equivalent Supervised Access format.

However, as can be seen from the distribution graphs below (and the variable 'Speed Corrected' in table 12.9), the Speed variable has some restriction in its range because a certain proportion of individuals attempt all questions on such tests and so achieve the maximum possible score on the Speed measure.

Figure 12.1 Swift Analysis Aptitude-R Speed Distribution (N=37,482)



This is where the Pace measure adds value in particular, as it can differentiate completion rates in individuals who have attempted all of the questions. That said, it must be borne in mind that Speed provides an advantage in itself as it can be manually scored by hand and so remains in the Hard-Copy versions of Saville Consulting's aptitude tests, as well as the equivalent online Supervised Access versions.

Once Speed has been corrected[1] for this inherent restriction of range it is reassuring that Pace and Speed are measuring very similar constructs, with correlations in the range of .88 to .93.

[1] *Analysis revealed the most suitable and effective correction to apply was based on a standard deviation four times that of the restricted Speed variable.*

Figure 12.2 Swift Analysis Aptitude-R Verbal Pace Distribution (N=27,852)

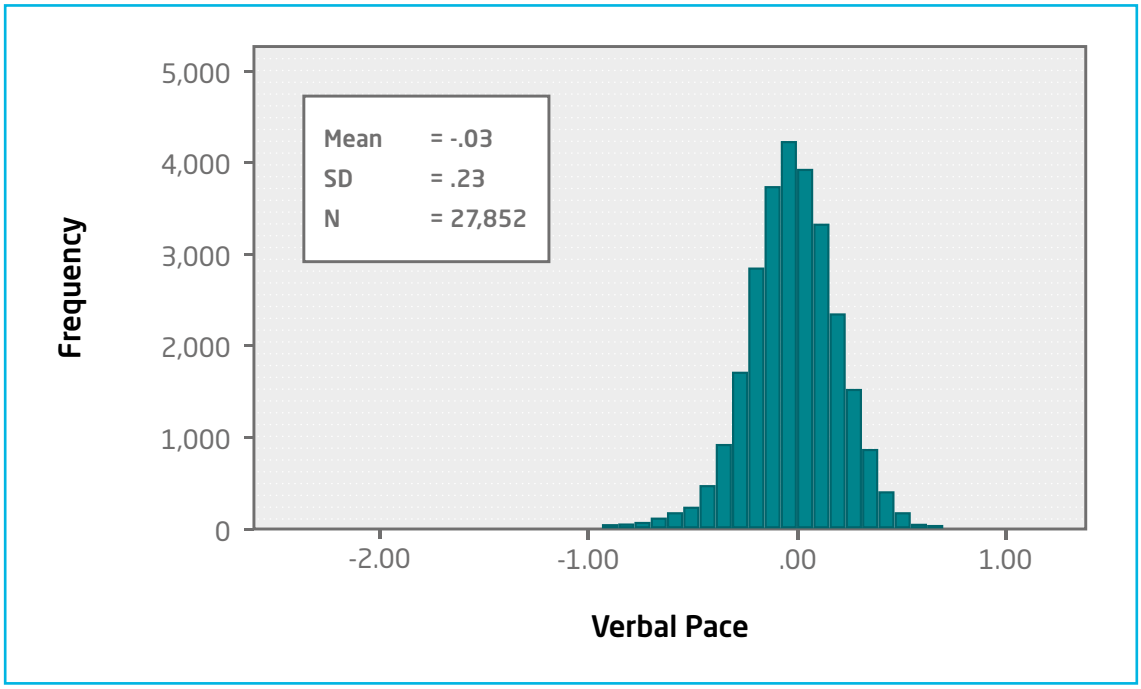


Figure 12.3 Swift Analysis Aptitude-R Numerical Pace Distribution (N=27,860)

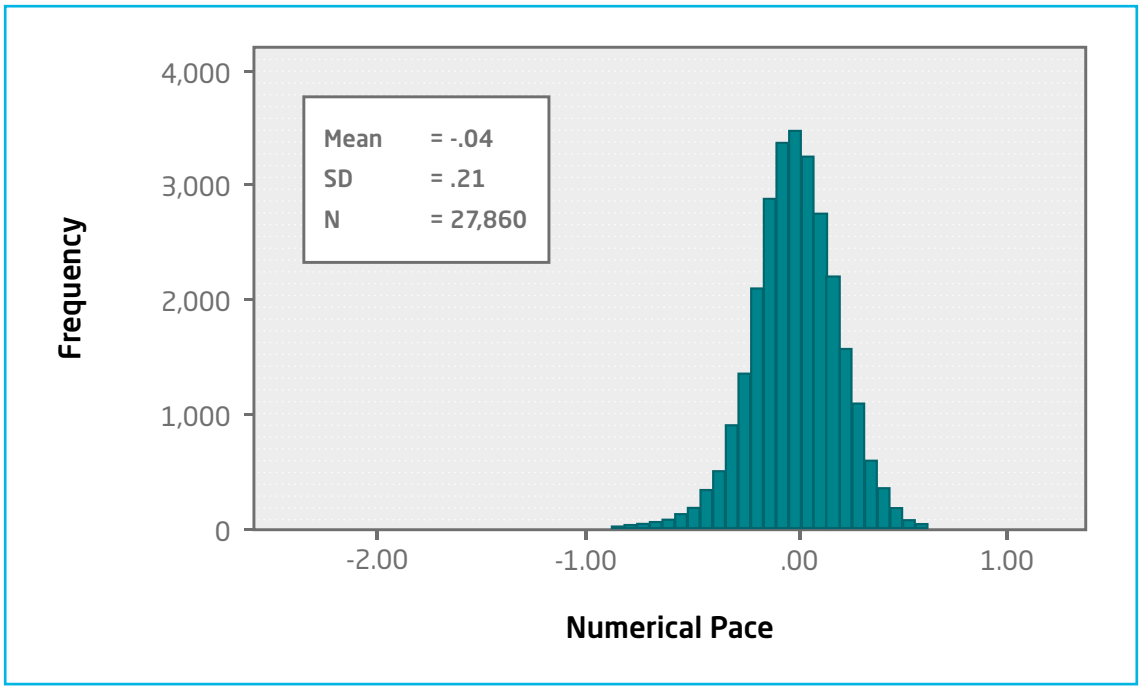
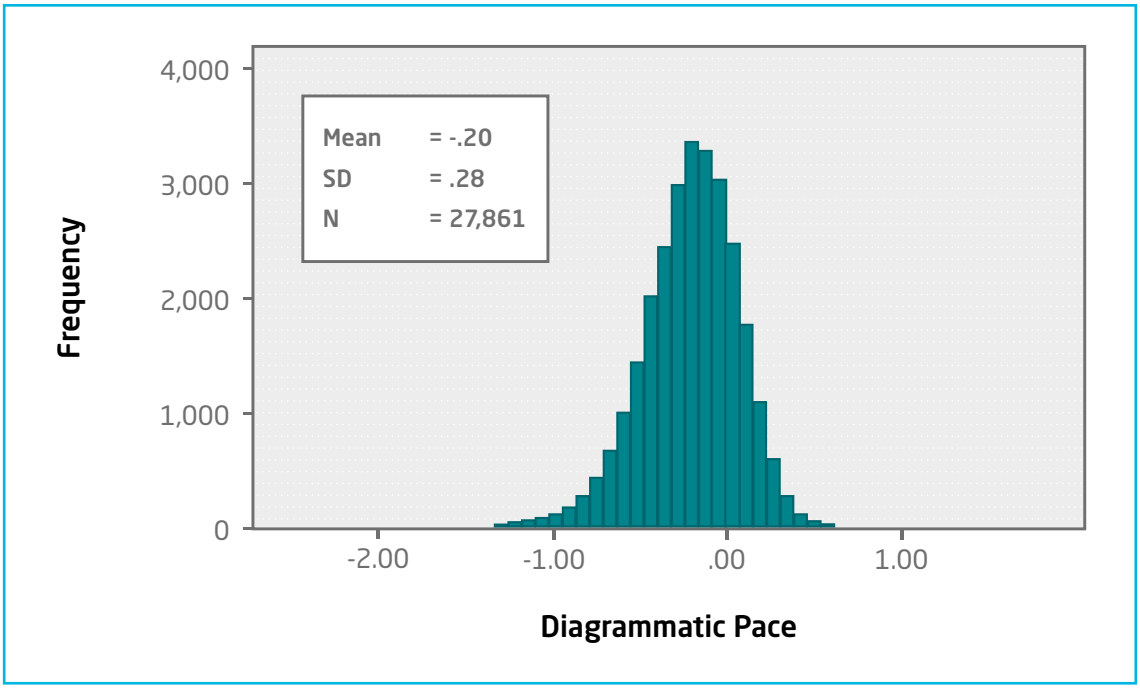


Figure 12.4 Swift Analysis Aptitude-R Diagrammatic Pace Distribution (N=27,861)



Intercorrelations of Swift Analysis Aptitude Sub-Scores

The Swift Analysis Aptitude overall and individual sub-test scores were intercorrelated in the Project Epsom sample, as shown below.

Table 12.11 Intercorrelations between Swift Analysis Aptitude Total Scores (N=308)

	Mean	SD	Total	Verbal	Numerical	Diagrammatic
Total	11.43	3.78		.78	.74	.68
Verbal	3.78	1.93			.36	.27
Numerical	4.73	1.62				.29
Diagrammatic	1.93	1.59				

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.

As can be seen, the intercorrelations between the total scores for the three sub-tests are moderate in strength (.36, .27, .29) which suggests that the different areas tap into different constructs as expected.

Intercorrelations of Professional Aptitude Total Scores

The Professional Aptitudes overall and individual test total scores were intercorrelated, as shown in the table below.

Table 12.12 Intercorrelations between Professional Aptitudes Scales (N=300)

	Mean	SD	Total Score	Verbal Total	Numerical Total	Diagrammatic Total
Overall Total	49.95	12.94		.87	.86	.80
Verbal Total	17.75	5.51			.61	.54
Numerical Total	13.93	5.36				.53
Diagrammatic Total	18.26	4.48				

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=300.

The tests within the Professional Aptitude series do not correlate too strongly, so appear to be measuring separate constructs. In this sample, the Numerical test appeared to be the most difficult, with the mean correct score being 13.93 out of 28. Overall speed, accuracy and caution scores were also derived and intercorrelated, as shown below.

Table 12.13 Intercorrelations between Professional Aptitudes Total Score and the Original Test-Taking Information (N=300)

	Mean	SD	Total Score	Accuracy	Speed	Caution
Overall Total	49.95	12.94		.82	62	.25
Accuracy	.72	.14			.08	.75
Speed	.82	.12				-.60
Caution	-.11	.18				

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=300.

As can be seen, accuracy and speed appear to be largely independent of each other, with accuracy contributing more to total scores than speed. Total scores achieved appear to be relatively independent of the degree of the caution exercised when completing the test.

Intercorrelations of Work Aptitude Total Scores

The Work Aptitudes overall and individual test total scores were intercorrelated, as shown below.

Table 12.14 Intercorrelations between Work Aptitudes Scales (N=339)

	Mean	SD	Total Score	Verbal Total	Numerical Total	Diagrammatic Total
Overall Total	52.18	12.17		.84	.86	.83
Verbal Total	19.08	4.90			.57	.55
Numerical Total	15.33	5.13				.59
Diagrammatic Total	17.78	4.36				

Note: Any raw correlation higher than .11 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .09 is statistically significant at the p<.05 level (one-tailed). N=339.

All three tests in the Work Aptitudes series show construct independence as they do not correlate too highly with each other. The Numerical test appears to be the most difficult, with the average total score being 15.33 out of a possible total of 28.

Overall speed, accuracy and caution scores were also derived and intercorrelated, as shown below.

Table 12.15 Intercorrelations between Work Aptitudes Test-Taking Style Scores (N=339)

	Mean	SD	Total Score	Accuracy	Speed	Caution
Overall Total	52.18	12.17		.83	.52	.32
Accuracy	.86	.11			-.03	.79
Speed	.72	.14				-.64
Caution	-.14	.18				

Note: Any raw correlation higher than .11 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .09 is statistically significant at the p<.05 level (one-tailed). N=339.

Accuracy and speed appear to be independent of each other and accuracy contributed more to the total score achieved than either the speed or caution exercised by the respondent.

Correlations between Swift Analysis Aptitude Scores and the Work Aptitudes Standardization

There are eight 'link items' which appear in both the Work Aptitudes and Swift Analysis Aptitude (hard-copy) tests (see 'Construction' chapter for further details). This means it is possible to simulate Swift Analysis Aptitude correlations by taking just the eight link items from the Work Aptitudes standardization group. The correlations between these eight link items and the rest of the Work Aptitudes assessment are shown below.

Table 12.16 Correlations between Swift Analysis Aptitude Hard-Copy (SAA) Link Items and the Work Aptitudes Tests (N=339)

	Work Aptitudes Total	Work Aptitudes Accuracy	Work Aptitudes Speed	Work Aptitudes Caution	Work Verbal Analysis	Work Numerical Analysis	Work Diagrammatic Analysis
Total	**.91**	.82	.37	.39	.76	.78	.76
Accuracy	.77	**.89**	.02	.67	.67	.66	.63
Speed	.50	.06	**.84**	-.47	.37	.45	.44
Caution	.39	.73	-.45	**.84**	.37	.32	.30
Verbal	.69	.64	.25	.33	**.80**	.48	.46
Numerical	.77	.65	.40	.25	.53	**.86**	.53
Diagrammatic	.66	.62	.21	.35	.45	.45	**.80**

Note: Any raw correlation higher than .11 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .09 is statistically significant at the p<.05 level (one-tailed). N=339.

Correlations between the eight link items and the rest of the Work Aptitudes assessments are high, suggesting that the link items suitably sample the same areas of aptitude as the longer Work Aptitudes assessments. All scores matched by type are high above the various and commonly used thresholds proposed for the construct equivalence or parallel forms of measurement.

Correlations between Swift Analysis Aptitude and Other Saville Consulting Aptitude Assessments

Individuals in Project Epsom (N=308) completed the Swift Technical (STA), Analysis (SAA) and Comprehension (SCA) Aptitude assessments. As part of the construct validation of the Swift Analysis Aptitude assessments, correlations were run between the scales of the SAA and the other Swift assessments.

The hypotheses here are relatively straightforward in that we would expect each test and sub-test to correlate more closely with similar aptitude areas in other tests than less related areas, e.g., we expect the verbal sub-score from Swift Analysis Aptitude to correlate most highly with other verbal tests.

Table 12.17 Correlations between Swift Analysis Aptitude and Swift Technical Aptitude Scores (N=308)

	STA Total	STA Spatial	STA Mechanical	STA Diagrammatic
SAA Total	**.53**	.21	.33	.54
SAA Verbal	.39	.16	.21	.43
SAA Numerical	.36	.12	.27	.35
SAA Diagrammatic	.42	.18	.27	.40

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 SAA and STA tests correlate relatively moderately with each other and so appear, as predicted, to be picking up unique aspects of the workplace aptitude construct space. This can be seen in that the Total Scores only correlate .53.

Table 12.18 Correlations between Swift Analysis Aptitude and Swift Comprehension Aptitude Scores (N=308)

	SCA Total	SCA Verbal	SCA Numerical	SCA Error Checking
SAA Total	**.54**	.43	.39	.43
SAA Verbal	.47	**.43**	.28	.37
SAA Numerical	.39	.30	**.36**	.25
SAA Diagrammatic	.32	.19	.22	.32

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 SAA and SCA tests do not correlate too highly and so do appear to be picking up different aspects of the workplace aptitude construct space. This can be seen in that the total scores only correlate .54. The two matched test pairs (SAA Verbal and SCA Verbal, SAA Numerical and SCA Numerical) correlate .43 and .36 respectively. It must be noted that the tests are designed to cover different levels of difficulty, so very high correlations would not be expected here.

Table 12.19 Correlations between Swift Analysis Aptitude and Practical Aptitudes Scores (N=103)

	Practical Aptitudes Overall Total	Practical Spatial Reasoning Total	Practical Mechanical Reasoning Total	Practical Diagrammatic Reasoning Total
SAA Total	**.65**	.44	.56	.62
SAA Verbal	.37	.22	.34	.36
SAA Numerical	.52	.41	.41	.46
SAA Diagrammatic	.54	.33	.47	**.53**

Note: Any raw correlation higher than .20 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .17 is statistically significant at the p<.05 level (one-tailed). N=103.

The overall correlation between SAA and the Practical Aptitudes is relatively high, at .65. Although they were designed to cover different levels of difficulty, the related SAA Diagrammatic and Practical Diagrammatic tests correlate .53.

Table 12.20 Correlations between Swift Analysis Aptitude and Operational Aptitudes Scores (N=103)

	Operational Aptitudes Overall Total	Operational Verbal Comprehension Total	Operational Numerical Comprehension Total	Operational Error Checking Total
SAA Total	.69	.60	.74	.62
SAA Verbal	.47	.54	.48	.25
SAA Numerical	.54	.41	.63	.37
SAA Diagrammatic	.50	.34	.51	.44

Note: Any raw correlation higher than .20 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .17 is statistically significant at the p<.05 level (one-tailed). N=103.

There is a relatively high level of correlation between the SAA and the Operational Aptitudes Scores (.69). The related SAA Numerical and Operational Numerical tests correlate .63, while the related SAA Verbal and Operational Verbal tests correlate a little less strongly at .54.

Correlations between Work Aptitudes and Other Aptitude Assessments which Measure Similar Reasoning Areas

A construct validation study was carried out using Work Aptitudes. 98 participants took part in this study, completing the Work Aptitudes alongside SHL tests which measure similar reasoning areas: 'Verbal Critical Reasoning' (VMG6), 'Numerical Critical Reasoning' (NMG6) and 'Diagramming' (DIT5). The design was counter-balanced with half of the participants completing the Work Aptitudes before the SHL tests and the other half completing the SHL tests first. The descriptive statistics and correlations are shown in the following tables.

Table 12.21 Total Score Descriptive Statistics for Work Aptitudes and Other Aptitude Assessments which Measure Similar Reasoning Areas (N=98)

Test	Number of Questions	Study Mean Score	Study SD	Study % Correct	Norm Mean Score	Norm SD
Verbal	28	18.29	4.39	64%	16.20	5.60
Numerical	28	13.72	5.10	50%	14.00	4.90
Diagrammatic	28	16.47	4.30	57%	17.53	4.51
SHL VMG6	32	19.14	4.60	59%	20.65	5.05
SHL NMG6	20	6.60	3.20	35%	12.16	4.18
SHL DIT5	50	28.43	7.50	56%	33.30	8.05

Note: Any raw correlation higher than .20 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .17 is statistically significant at the p<.05 level (one-tailed). N=98.

The mean total scores for five of the six tests were in line with published norms. However, scores on NMG6 were more than one standard deviation below the published SHL norms. In fact, while the published SHL norm presents a mean score of around 12, our sample only managed to attempt 12 questions on average. Item analysis revealed NMG6 to be of a high level of difficulty, resulting in an average score of just 6.60 on a 20-item test. Contrastingly, the Numerical Analysis test in the Work Aptitudes range has a mean of 13.72 on a 28-item test.

Table 12.22 Correlations between Work Aptitudes and SHL Aptitude Tests (N=98)

	SHL Overall Total	VMG6 Total	NMG6 Total	DIT6 Total
Overall Total	**.79**	.59	.68	.65
Verbal Total	.60	**.60**	.47	.37
Numerical Total	.71	.48	**.63**	.59
Diagrammatic Total	.67	.37	.58	**.65**

Note: Any raw correlation higher than .20 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .17 is statistically significant at the p<.05 level (one-tailed). N=98.

All correlations for the Work Aptitudes total scores with scores on similar SHL tests reached at least .60, supporting the construct validity of the Work Aptitudes assessments.

Correlations between Professional Verbal Analysis, Professional Numerical Analysis and Raven's Advanced Progressive Matrices

58 individuals in a Police Superintendent selection process completed the Professional Verbal Analysis, Professional Numerical Analysis and Raven's Advanced Progressive Matrices tests. These correlations are shown below.

Table 12.23 Correlations between Professional Verbal Analysis, Professional Numerical Analysis and Raven's Advanced Progressive Matrices (N=58)

	Professional Verbal Analysis	Professional Numerical Analysis	Raven's Advanced Progressive Matrices
Professional Verbal Analysis		.43	.50
Professional Numerical Analysis			.46
Raven's Advanced Progressive Matrices			

Note: Any raw correlation higher than .26 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .22 is statistically significant at the p<.05 level (one-tailed). N=58

Professional Verbal Analysis and Professional Numerical Analysis correlated .50 and .46 respectively with the total score on Raven's Advanced Progressive Matrices. The two Professional Aptitude tests correlated .43 with each other.

Correlations between Professional Verbal Analysis, Professional Numerical Analysis and the Watson-Glaser Critical Thinking Appraisal

70 senior managers in the construction industry completed the Professional Verbal Analysis, Professional Numerical Analysis and Watson-Glaser Critical Thinking Appraisal tests. These correlations are shown below.

Table 12.24 Correlations between Professional Verbal Analysis, Professional Numerical Analysis and the Watson-Glaser Critical Thinking Appraisal (N=70)

	Professional Verbal Analysis	Professional Numerical Analysis	Watson-Glaser Critical Thinking Appraisal
Professional Verbal Analysis		.50	.61
Professional Numerical Analysis			.46
Watson-Glaser			

Professional Verbal Analysis and Professional Numerical Analysis correlated .61 and .46 respectively with the total score on the Watson-Glaser Critical Thinking Appraisal assessment. The two Professional Aptitude tests correlated .50 with each other.

12.6 Factor Structure Analyses

Defining Effectiveness at Work using Aptitude Assessments and Effectiveness Ratings

In order to establish and compare the hierarchical structure of aptitude (predictor) and aligned ability (criteria) scales, factor analyses were carried out across three Saville Consulting Swift Assessments, across the six Wave Performance 360 ability self-ratings and across the six Wave Performance 360 ability external ratings. The more variance in effectiveness that can be explained by the scores on these questionnaires, the more confident we can be that the assessments are valid measures of workplace effectiveness.

Table 12.25 Principal Components Analyses (Unrotated) Derived from Swift (SAA, SCA and STA), Performance 360 (Self) and Performance 360 (Rater) Variables (N=308)

Swift Predictor Scales	Swift Scales		Performance 360 Criteria	Self Ratings		External Ratings	
PCA Component	1st	2nd	PCA Component	Self 1st	Self 2nd	Rater 1st	Rater 2nd
Variance Explained	38%	11%	Variance Explained	52%	19%	50%	52%
Verbal Analysis	.66	-.30	Working with Words	.37	-.78	.46	-.72
Verbal Comprehension	.68	-.36					
Numerical Analysis	.60	-.03	Working with Numbers	.81	-.07	.82	-.22
Numerical Comprehension	.60	.01					
Error Checking	.70	-.21	Working with Details	.70	-.40	.77	-.35
Diagrammatic Analysis	.56	.25	Working with Systems	.82	.00	.79	.02
Diagrammatic Reasoning	.73	-.11					
Spatial Reasoning	.43	.54	Working with Designs	.77	.39	.71	.52
Mechanical Reasoning	.49	.62	Working with Equipment	.75	.45	.65	.59

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.

Table 12.25 shows the variable loadings on the 1st and 2nd statistical components when entering the three sub-scores of each of the three Swift aptitude series', six Performance 360 Self criteria or six Performance 360 Rater criteria into a principal components analysis (PCA).

Each PCA produced two factors with Eigenvalues >1. For the aptitude assessment sub-tests the 'General Reasoning' loadings on the first component are highest for 'Diagrammatic Reasoning' followed by 'Error Checking' and 'Verbal Comprehension'. 'Spatial Reasoning' and 'Mechanical Reasoning' have the lowest component loadings of all of the aptitude assessments.

For self-ratings as well as external ratings Working with Systems and Working with Numbers had the highest loadings, with lower loadings for Working with Words.

The second component in all three analyses resembles the classic 'Verbal/Educational' versus 'Spatial/Mechanical' distinction of Vernon (1950). The Diagrammatic/Working with Systems and Numerical/Working with Numbers concepts are at the core of the General Reasoning Ability research (e.g., Kurz, 2000).

Professional Aptitudes: Factor Analysis of Sub-Scores

Statistically independent accuracy, speed and item type sub-scores were entered into a principal components factor analysis followed by Varimax rotation resulting in four components with Eigenvalues >1.

The unrotated solution shown in table 12.26 reveals high 'g' loadings for virtually all sub-scores. The first three rotated components produce verbal, numerical and diagrammatic factors with homogeneous loadings of the various item type sub-scores, while the fourth is a test-taking style factor which contrasts speed with accuracy.

Table 12.26 Professional Aptitudes Sub-Scores Factor Analysis (N=266)

	Component Matrix				Rotated Component Matrix			
	1	2	3	4	1	2	3	4
Verbal								
Accuracy	**.75**	**-.52**			**.86**	.21	.23	
Speed	**.61**			**.50**	**.53**	.24		**.56**
Understanding Word Meaning	**.56**	-.32			**.64**			
Comprehending Text	**.71**			.31	**.71**	.30		
Making Verbal Inferences	**.63**	-.37			**.68**		.21	
Evaluating Written Materials	**.73**	-.21		.38	**.78**	.22		
Comparing Arguments	**.74**	-.26		.31	**.78**	.22	.22	
Numerical								
Accuracy	**.66**	-.39		-.42	**.41**	**.61**	.20	**-.44**
Speed	**.49**	**.59**	**-.42**			**.59**		**.64**
Understanding Tables	**.77**		-.24	-.22	.35	**.71**	.27	
Comprehending Graphs	**.65**		-.39		.26	**.73**		
Making Numerical Inferences	**.67**		-.35	-.20	.20	**.74**		
Evaluating Quantities	**.70**		-.33	-.26	.21	**.76**	.23	
Comparing Data	**.68**		**-.42**	-.21	.22	**.79**		
Diagrammatic								
Accuracy	**.63**		**.50**	**-.42**	.30	.21	**.74**	**-.42**
Speed	.35	**.69**		.36			**.41**	**.76**
Understanding Logic Rules	**.62**	.22	**.45**		.26		**.73**	
Comprehending Process Diagrams	**.48**	**.42**	**.41**				**.73**	
Identifying Causes	**.59**	.21	**.42**		.25		**.69**	
Finding Faults	**.65**		**.44**		.27		**.74**	
Comparing Flowchart Sequences	**.60**	.21	.22		.20	.29	**.57**	

Note: Any raw correlation higher than .13 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .11 is statistically significant at the p<.05 level (one-tailed). N=266

Work Aptitudes: Factor Analysis of Sub-Scores

Statistically independent accuracy, speed anditem type sub-scores were entered into a principal components factor analysis followed by Varimax rotation resulting in four components with Eigenvalues >1.

The unrotated solution shown in table 12.27 reveals high 'g' loadings for virtually all sub-scores. The first three rotated components produce verbal, numerical and diagrammatic factors with homogenous loadings of the various item type sub-scores, while the fourth is a test-taking style factor which contrasts speed with accuracy.

Table 12.27 Work Aptitudes Sub-Scores Factor Analysis (N=339)

	Component Matrix				Rotated Component Matrix			
	1	2	3	4	1	2	3	4
Verbal								
Accuracy	**.76**	-.26	**.48**		.25	**.87**	.24	
Speed	**.46**	**.57**	.33			**.47**		**.64**
Understanding Word Meaning	**.57**		**.44**			**.71**		
Comprehending Text	**.59**		.31		.23	**.62**	.20	
Making Verbal Inferences	**.56**		**.42**			**.66**		.23
Evaluating Written Materials	**.66**		.36			**.66**	.30	
Comparing Arguments	**.70**		**.42**		.20	**.74**	.22	.23
Numerical								
Accuracy	**.70**	**-.48**	-.23	-.24	**.68**	.31	.33	**-.41**
Speed	**.42**	**.69**	-.27	-.24	**.54**			**.70**
Understanding Tables	**.75**		-.26	-.26	**.74**	.29	.28	
Comprehending Graphs	**.61**	.22	-.21	-.33	**.67**	.24		.26
Making Numerical Inferences	**.60**		-.32	-.29	**.70**			.23
Evaluating Quantities	**.71**		-.31	-.23	**.73**	.22	.28	
Comparing Data	**.69**		**-.33**	-.26	**.74**		.25	
Diagrammatic								
Accuracy	**.65**	-.56		.33	.28	.24	**.75**	**-.41**
Speed	.35	**.75**		.34			.32	**.83**
Understanding Logic Rules	**.61**	-.29		.21	.30	.25	**.58**	
Comprehending Process Diagrams	**.61**			**.46**			**.71**	
Identifying Causes	**.55**			**.50**			**.71**	
Finding Faults	**.66**			**.46**		.27	**.73**	
Comparing Flowchart Sequences	**.62**		-.26	.32	.34		**.61**	.32

Note: Any raw correlation higher than .11 is statistically significant at the p<.05 level (two-tailed) and any raw correlation higher than .09 is statistically significant at the p<.05 level (one-tailed). N=339

12.7 A Summary of the Validity of the Analysis Aptitude Range

Criterion-Related Validity

The table below summarizes key evidence of the criterion-related validity of the Analysis Aptitude Range.

Table 12.28 A Summary of the Criterion-Related Validity of the Analysis Aptitude Range

Analysis Aptitude Test	Correlate Used	Sample Size	Correlation (Uncorrected)	Correlation (Corrected)
Validity of Swift Analysis Aptitude with Ability Criteria				
SAA Total	Sum of Scores on Working with Words, Numbers and Systems Competencies	308	.29	.54
SAA Verbal	Score on Working with Words Competency	308	.27	.48
SAA Numerical	Score on Working with Numbers Competency	308	.20	.34
SAA Diagrammatic	Score on Working with Systems Competency	308	.10	.24
Validity of Swift Analysis Aptitude with Various Other Criteria				
SAA Total	Accountancy Exam - Mean Correlation Across 9 Modules	178	.24	.31
SAA Total	Selection Centre Score - Group Exercise	188	.28	.36
SAA Total	Selection Centre Score - Written Exercise	188	.23	.30
SAA Total	Selection Centre Score - Simulation Exercise	188	.32	.41

Note: The correlations between Swift Analysis Aptitude and the Ability Criteria were adjusted for criterion unreliability based on the inter-rater reliability of peer ratings on these criteria (Overall Sum of Competencies = .29, Working with Words =.31, Working with Numbers = .34, Working with Systems = .18). N=263. The correlations between Swift Analysis Aptitude and the other criteria were adjusted according to an attenuation estimate of .60.

Construct Validity

The table below summarizes key evidence of the construct validity of the Analysis Aptitude Range.

Table 12.29 A Summary of the Construct Validity of the Analysis Aptitude Range

Analysis Aptitude Test	Correlate Used	Sample Size	Correlation (Uncorrected)
Validity of Swift Analysis Aptitude with other Ability Tests			
SAA Total Score	Swift Comprehension Aptitude (IA) - Total Score	308	.54
	Swift Technical Aptitude (IA) - Total Score	308	.53
	Work Aptitudes (HC) - Overall (Combined) Total Score	339	.91
	Operational Aptitudes (HC) - Overall (Combined) Total Score	103	.69
	Practical Aptitudes (HC) - Overall (Combined)Total Score	103	.65
Validity of Professional Aptitudes with other Ability Tests			
Professional Verbal Analysis - Total Score	Score on Raven's Advanced Progressive Matrices	58	.50
Professional Numerical Analysis - Total Score		58	.46
Professional Verbal Analysis - Total Score	Score on Watson-Glaser Critical Thinking Appraisal	70	.61
Professional Numerical Analysis - Total Score		70	.46
Validity of Work Aptitudes with other Ability Tests			
Work Aptitudes - Overall (Combined) Total Score	SHL Aptitude Test - Combined Verbal, Numerical and Diagrammatic Score	98	.79
Work Verbal Analysis - Total Score	SHL Aptitude Test - Verbal Score	98	.60
Work Numerical Analysis - Total Score	SHL Aptitude Test - Numerical Score	98	.63
Work Diagrammatic Analysis - Total Score	SHL Aptitude Test - Diagrammatic Score	98	.65
Validity of Professional Aptitudes with GCSE Scores			
Professional Aptitudes - Overall (Combined) Total Score	UK GCSE English, Mathematics and Science Combined Score	227	.53
Validity of Work Aptitudes with GCSE Scores			
Work Aptitudes - Overall (Combined) Total Score	UK GCSE English, Mathematics and Science Combined Score	273	.62

Summary

This chapter has provided validity evidence for both the Swift combined and in-depth single tests in the Analysis Aptitude Range.

Swift Analysis Aptitude achieves impressive levels of criterion-related validity for an assessment that can be completed in 18 minutes. Given that it samples the same domains as the in-depth single tests, it is appropriate to assume that Swift Analysis Aptitude provides a conservative lower-bound estimate of the validity of the longer tests within the Analysis Aptitude Range.

The tests within the Analysis Aptitude Range show appropriate and expected correlations with other test scores. The level of correlations between total scores of other tests suggests that the broad domains of performance being assessed have some overlap as expected, but equally provide a good degree of unique performance measurement. The matched correlations between the sub-test areas (e.g., Verbal Analysis in Swift Analysis Aptitude and Verbal Comprehension in Swift Comprehension Aptitude) are reflective of the expected level of construct overlap between these sub-tests. The relatively moderate correlations between the different sub-areas within the Analysis Aptitude Range tests are also indicative of their measuring sufficiently different aspects of workplace performance.

Appendix C details supplementary validity evidence for some of the test-taking information scores in the Analysis Aptitude Range.

Appendix D summarizes correlations between Swift Analysis Aptitude and personality questionnaires featured in Project Epsom.

Appendix E includes full correlation matrices showing the relationship between scores on Swift Analysis Aptitude and a range of other assessments featured in Project Epsom.