Validation of the Saville Consulting Analysis Aptitude Assessment Range

Kurz, R., Saville, P., MacIver, R., Mitchener, A., Parry, G., Oxley, H. & Archer, J.; Saville Consulting Group; Lishman, J., Independent Consultant; Mason, J., Scott Jardine; Roscorla, P., Home Office

Summary

This paper describes the development and validation of a new range of Analysis Aptitude Assessments developed by Saville Consulting. It presents reliability and validity figures from the standardisation of Professional and Work Aptitudes as well as data on the short Swift Analysis Aptitude measure. Construct validity evidence is presented against other aptitude tests as well as against NEO and Saville Consulting Wave self-report questionnaires.

Introduction

Recent meta-analytical research studies have confirmed that aptitude tests are the most valid predictors of occupational performance e.g. Bertua, Anderson & Salgado (2005) found 'operational validities' in the magnitude of 0.5–0.6. In spite of this, a commonly voiced concern is that ability testing has stood still, and that many measures have become dated.

The purpose of this paper is to outline the validity of the new Aptitude Assessment range developed by Saville Consulting. The portfolio includes online and hard copy tests and offers users the flexibility of both unsupervised and supervised parallel versions. Each test provides additional scores to give test user, candidate and line manager greater insight into test performance.

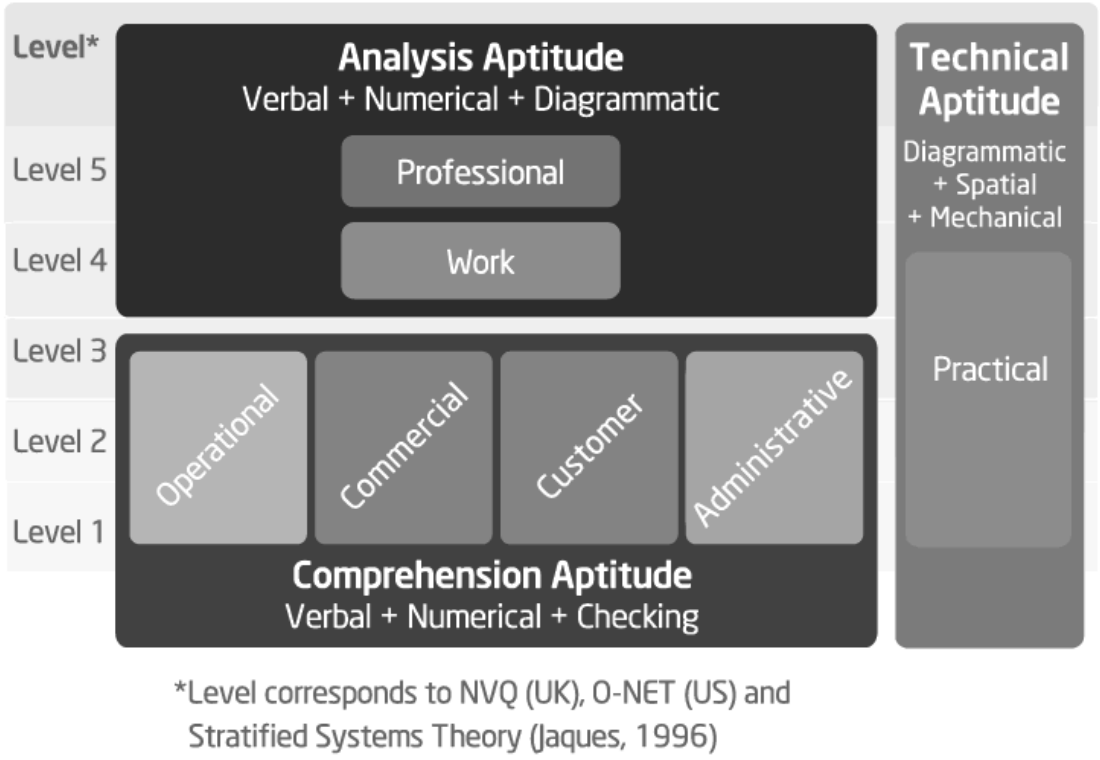
The paper outlines innovative security and user friendly features and provides data on the psychometric quality indicators of the Analysis Aptitude range of Verbal, Numerical and Diagrammatic Analysis tests.

Aptitude Assessment Portfolio

Figure 1 outlines the structure of the new Saville Consulting Aptitude Assessment offering. Test ranges have been designed for specific complexity levels as defined through the National Vocational Qualifications (NVQ) framework and Jaques (1996) Stratified Systems Theory. There are three major ranges which feature three test formats each.

The Analysis Aptitude range features high-level Verbal, Numerical and Diagrammatic Analysis tests for complex roles (NVQ Levels 4-5). The Swift Analysis Aptitude combination assessment covers each ability area through short sub-tests (6 minutes for 8 questions) with a total testing time of 18 minutes for 24 questions. There are also two parallel series of individual tests to separately measure Verbal, Numerical and Diagrammatic Analysis (20 minutes for 28 questions each). Professional Aptitudes is designed for experienced professionals, managers and directors, and Work Aptitudes for graduates and technicians.

Figure 1: Saville Consulting Aptitude Assessment Portfolio



The Comprehension Aptitude range features Verbal Comprehension, Numerical Comprehension and Error Checking tests for general roles (NVQ Levels 1-3) with four parallel series that are broadly aligned to one of the four Saville Consulting Wave clusters (MacIver et al., 2006), and corresponding job roles.

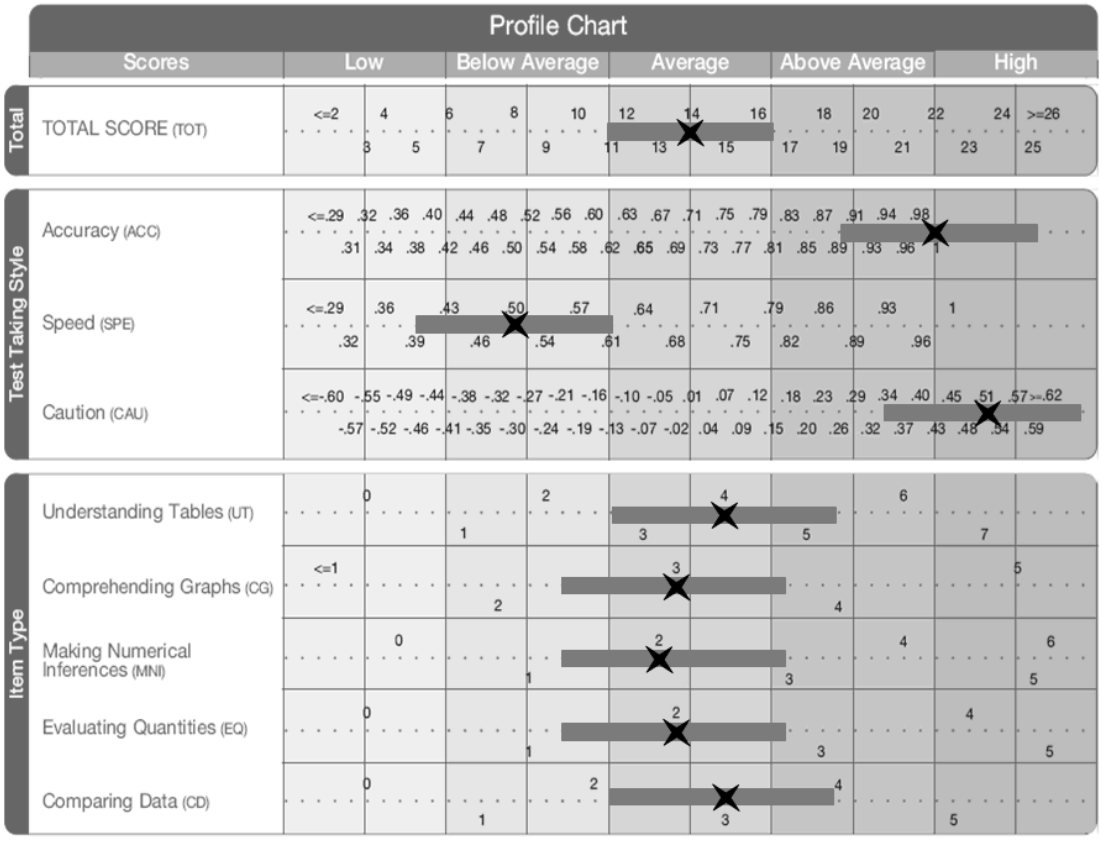
The Technical Aptitude range features Diagrammatic, Spatial & Mechanical Reasoning tests for production, construction, engineering and scientific roles.

Test Taking Style Scores

Aptitude tests have successfully been used over the last century largely based on total scores, i.e. the number of questions answered correctly. The Saville Consulting Aptitude Assessments challenge this orthodoxy as candidates of equal ability level vary widely in their

speed-accuracy balance when taking tests (Kurz, 2000). Normed Test Taking Style sub-scores have been developed for all Saville Consulting Aptitude Assessments to measure accuracy, speed and caution in test completion. Figure 2 shows the profile of 'Cathy Cautious' on Professional Numerical Analysis illustrating these sub-scores in practice. Cathy has answered 14 of 28 questions, and got all of them correct resulting in an 'Average' score. She made no mistakes so her Accuracy (proportion correct) is high. She only attempted half of the test questions resulting in 'Below Average' Speed (proportion of test questions attempted). She chose to prioritise accuracy over speed and consequently obtained a 'High' score on Caution. The behaviour in the test taking situation informs the feedback session, and allows predictions to be made about likely work performance i.e. balance of quantity and quality of work.

Figure 2: Professional Numerical Analysis Profile Chart for 'Cathy Cautious'



Item Type Scores

Each test in Professional and Work Aptitudes features five distinct item formats that result in normed Item Type sub-scores. For Cathy all item type scores are in the 'Average' range – in this situation the sub-score names simply aid interpretation and feedback of the overall 'Numerical Analysis' construct measured. Any score differences observed across the item

types highlight the type of questions candidates found particularly easy or problematic, and highlight relative strengths as well as potential development needs.

Overall Scores

The co-standardisation of Professional Verbal, Numerical and Diagrammatic Analysis allows the summation and norming of scores across the three measures. The resulting 'Professional Aptitude' or 'Work Aptitude' Overall Score represents a clearly defined higher-order construct of analytical reasoning that ties down abstract conceptions of 'g', intelligence or general mental ability.

Overall Scores across several tests are ideal for assessment situations where a single rating on a dimension such as 'Analytical Thinking' is required, and to capture the combined validity of a series of tests.

We have utilised the principle of higher-order 'Overall Scores' to create Swift Analysis Aptitude combination measures that are based on short (6 minutes) Verbal, Numerical and Diagrammatic Analysis sub-tests. Invited Access online versions are ideal for developmental audits and for screening out unsuitable applicants through unsupervised online testing early on in the recruitment process. Supervised Access online and Hard Copy versions offer secure verification of results, and/or can be deployed in assessment situations where time is at a premium e.g. in Assessment Centres and prior to interviews.

Method

A series of studies was conducted to ascertain the reliability and validity of the new measures. The main study reported here is based on the standardisation of Professional Aptitudes on a total of N=300 professionals and graduates who completed hard copy versions of Verbal, Numerical and Diagrammatic Analysis during the 2005 standardisation. Sub-samples completed a Test-Retest study (N=120) or Alternate Form reliability studies where individuals also completed Work Aptitudes parallel tests (N=109 for Verbal and Numerical; N=135 for Diagrammatic). Complete sets of educational criteria (GCSE results) and Reasoning Competency self-efficacy ratings were available for N=227 participants. A number of other studies and their results are briefly outlined below (without reference to the various Item Type sub-scores that tend to show patterns similar to Total Scores).

Reliability

Reliabilities for the three Professional Aptitude tests are given in Table 1. Internal Consistency reliabilities in the standardisation group were strong with Test-Retest and

Alternate Form reliabilities also reasonable despite of the relatively short nature of the tests. Item Type sub-scores had an average internal consistency of .47 and their average alternate form and test retest reliabilities were both .48, indicating that these scores should be interpreted in the light of the Total Score rather than in isolation. Test Taking Style sub-scores had an average test retest reliability of .54 and average alternate form reliability of .61 showing that there is a considerable degree of consistency in individuals speed-accuracy balance.

Table 1: Professional Aptitudes Reliabilities

	r Internal Consistency Reliability (N=300)	r Alternate Form Reliability (N=109-135)	r Test Retest Reliability (N=120)
Professional Verbal Analysis	.84	.71	.76
Professional Numerical Analysis	.87	.70	.72
Professional Diagrammatic Analysis	.80	.69	.71

Swift Analysis Aptitude is a short assessment consisting of separately timed Verbal, Numerical and Diagrammatic sub-tests of 6 minutes each. The Internal Consistency reliabilities for the Total Score of the supervised and unsupervised version were satisfactorily high at .78 (N=639) and .76 (N=275) respectively. Verbal, Numerical and Diagrammatic sub-scores achieved Internal Consistency reliabilities of .56, 69 and .61 for the supervised version, and .62, .70 and .44 for the unsupervised version.

In a field study N=183 Research Officer applicants completed Swift Analysis Aptitude (Invited Access) in advance of a hard copy verification session. Candidates were fully aware that they would be retested with parallel versions at the hard copy session and assumed the online tests would be used for screening out unsuitable candidates. In actual fact the unsupervised tests were used 'blind' (i.e. results ignored) to ascertain the correlation across mode of presentation in this screening & verification 2 stage format. Difference scores showed a symmetrical distribution with 2 individuals showing score improvements of 4 Stens, and 2 individuals showing score decrements of 4 Stens. 2 suspected 'Cheats' showed performance decrements of 5 and 6 Stens respectively and were removed from the reliability analysis. High Alternate Form reliabilities were achieved between Online Unsupervised and Hard Copy Supervised scores: Total Score .72 (.69 before removal of 'Cheats'), Accuracy .66, Speed .38, Caution .47, Verbal .51, Numerical .59, and Diagrammatic .36.

Criterion-related Validity

Table 2 shows validities for the three Professional Aptitudes tests and Overall scores (sum across Verbal, Numerical and Diagrammatic) against educational and competency criteria

unadjusted for any artefacts such as criterion unreliability or restriction of range. Hypothesised Overall Score and differential patterns (highlighted in bold) are largely supported by the data. The Overall Total Score exceeds the validity of any single test in the prediction of GCSE Results (sum of English, Mathematics and Science) while for Reasoning Competencies (sum of Working with Words, Numbers and Diagrams self-rating) Professional Numerical Analysis shows the highest validity. Of the Test Taking Style sub-score Overall Accuracy tends to account for most of the validity of tests followed by Speed and then Caution.

Table 2: Professional Aptitudes Validities (Unadjusted) against Educational and Competency Self-rating Criteria N=227

	GCSE Results	English	Mathematics	Science	**Reasoning Competencies**	Working with Words	Working with Numbers	Working with Diagrams
Professional Verbal Analysis	**.50**	**.47**	.39	.37	**.32**	**.34**	.19	.19
Professional Numerical Analysis	**.46**	.28	**.46**	.31	**.46**	.19	**.49**	.36
Professional Diagrammatic Analysis	**.39**	.16	.42	**.35**	**.23**	-.01	.24	**.30**
Overall Total Score	**.55**	**.37**	**.52**	**.42**	**.41**	**.22**	**.38**	**.35**
Overall Accuracy	**.49**	.29	.49	.38	**.35**	.25	.29	.26
Overall Speed	**.29**	.26	.23	.20	**.23**	.06	.24	.22
Overall Caution	**.21**	.08	.24	.18	**.13**	.16	.07	.06

In an organisational study a total of N=50 senior managers in a global Engineering company completed Swift Analysis Aptitude (Invited Access) and the Professional Styles instrument of the Saville Consulting Wave range. Line manager ratings of Overall Job Performance were significantly predicted by Swift Analysis Aptitude Total (.26), Accuracy (.24), and Diagrammatic (.28) scores.

Construct Validation

Lishman (2005) conducted a classic construct validity study on N=98 students and professionals with counterbalanced order design. Professional Verbal, Numerical and Diagrammatic Analysis correlated satisfactorily with comparable SHL tests VMG6 (.60), NMG6 (.63) and DIT5 (.65).

In a Police Superintendents selection process (N=58) Professional Verbal Analysis correlated .50 and Professional Numerical Analysis .46 with Raven's Advanced Progressive Matrices, and .43 with each other. In the Professional Aptitudes standardisation (N=300) Verbal correlated .54 and Numerical .53 with Diagrammatic, and .61 with each other.

A sub-set (N=171) of the Research Officers mentioned above also completed a knowledge test. The score on this test correlated significantly with Total Score (.57/.56), Accuracy (.56/.51), Speed (.18/.21), Caution (.46/.29), Verbal (.47/.52), Numerical (.54/50) and Diagrammatic (.22/.24) sub-scores on the Online Unsupervised and Hard Copy Supervised tests respectively.

In the course of the development of a 'Multicultural Professionals & Graduates' norm (N=281) of non-native speakers, students (N=199) at the University of the West Indies in Barbados completed Professional Aptitudes. The group performed almost 1 SD lower than the standardisation group on Verbal Analysis but only 0.4 SD lower on Numerical and Diagrammatic Analysis. A sub-group (N=113) also completed the Professional Styles questionnaire. A number of Professional Styles dimensions correlated significantly with aptitude scores. Rational correlated highly significantly (around .30) with Total Score, Accuracy and Caution scores on Numerical and Diagrammatic Analysis. Inventive correlated significantly (around .25) with Numerical and Diagrammatic Accuracy. Organised and Reliable correlated significantly negatively (around -.30) with Total Score, Accuracy and Caution on Verbal and Numerical Analysis.

Senior Civil Servants (N=100) completed Professional Verbal and Numerical Analysis alongside the NEO PIR 'Big 5' personality questionnaire. Scores were about 0.6 SD's above the Professionals & Graduates norm for Verbal, and 0.2 SD's for Numerical. NEO Neuroticism was significantly correlated with Verbal Accuracy (.19) and Caution (.17) and Numerical Total Score (.23) and Accuracy (.20). NEO Extraversion was significantly negatively related with Verbal Total Scores (-.21), Accuracy (-.34) and Caution (-.36). NEO Agreeableness correlated significantly negatively with Numerical Total Scores (-.18) and Speed (-.17). NEO Conscientiousness correlated significantly negatively with Verbal Accuracy (-.27) and Caution (-.34). NEO Openness showed no significant correlations with any test scores.

Conclusions

The paper demonstrates that breadth and depth of aptitude assessment can be enhanced through innovation. The Total Scores of all aptitude assessments researched have excellent reliability and validity, and their Test Taking Style and Item Type sub-scores contribute towards a better understanding of people and jobs. Invited Access, Supervised Access and Hard Copy parallel versions of tests give clients flexibility and choice on how to conduct their assessments, and provide the security and comfort of supervised tests that counter cheating particularly in highly competitive selection settings. The relationships of Total Score and Test Taking Style variables with personality constructs seem rather complex so that further research across multiple samples is required before any real conclusions can be drawn.

References

Bertua, C., Anderson, N. & Salgado, J. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, Volume 78, Number 3, September 2005, pp. 387-409.

Jaques, E. (1996). *Requisite Organization: 2nd Edition*. Virginia: Cason Hall.

Kurz, R. (2000). *The Facets of Occupational Testing: General Reasoning Ability, Residual Aptitudes & Speed-Accuracy Balance*. Unpublished PhD dissertation. Manchester: UMIST.

Kurz, R. (2006). *Test Taking Style & Item Type Scores in the Assessment of Professional Aptitudes.* BPS Occupational Psychology Conference. Leicester: BPS.

Kurz, R. (2006). *The Validity of Item Type, Test Taking Style, General and Profile Facet Scores in Professional Aptitude Assessments.* Work Psychology Congress in Leipzig.

Lishman, J. (2005). *A Construct Validation Study of a New Ability Test Battery.* Northumbria University: Unpublished MSc Thesis.

MacIver, R., Saville, P., Kurz, R., Mitchener, A., Mariscal, K., Parry, G., Becker, S., Saville, W., O'Connor, K., Patterson, R. & Oxley, H. (2006). Making Waves – Saville Consulting Wave Styles questionnaires. *Selection & Development Review*, Volume 22, No.2, April 2006.