



Saville Consulting Wave® Performance 360 Handbook

PART 3: TECHNICAL

This handbook has been generated electronically. Saville Consulting do not guarantee that it has not been changed or edited. We can accept no liability for the consequences of the use of this handbook, howsoever arising.



10.0 Construction

This chapter describes the development process used to construct Wave® Performance 360. A general background to Wave Performance 360 can be found in the 'Introduction' chapter of this volume.

10.1 Development Background

Development Goal

Saville Consulting Wave Performance 360 was developed as part of the Saville Consulting Wave suite of tools. The primary goal of the assessment is to provide different perspectives on core aspects of individuals' work performance that underpin effectiveness at work. The content of Wave Performance 360 is specifically designed to be work-relevant and to focus on attributes which underpin behavioral, ability and overall effectiveness at work.

The quest for enhanced validity is regarded as the central aim of any reputable test or questionnaire developer, and it is a key priority for assessment and test users. Users have the potential to maximize the benefit of assessments by using tools where criterion-related validity is higher. For the Wave suite and Performance 360, high validity is of prime importance.

Validation was fundamental to the development of the Wave Performance 360 questionnaire. Section 15.2 'Wave Performance 360 and Superordinate Constructs' in the 'Validity' chapter (Chapter 15) provides a priori evidence post standardization that effectively cross-validates the behaviors assessed in Wave Performance 360 as underpinning overall effectiveness at work.

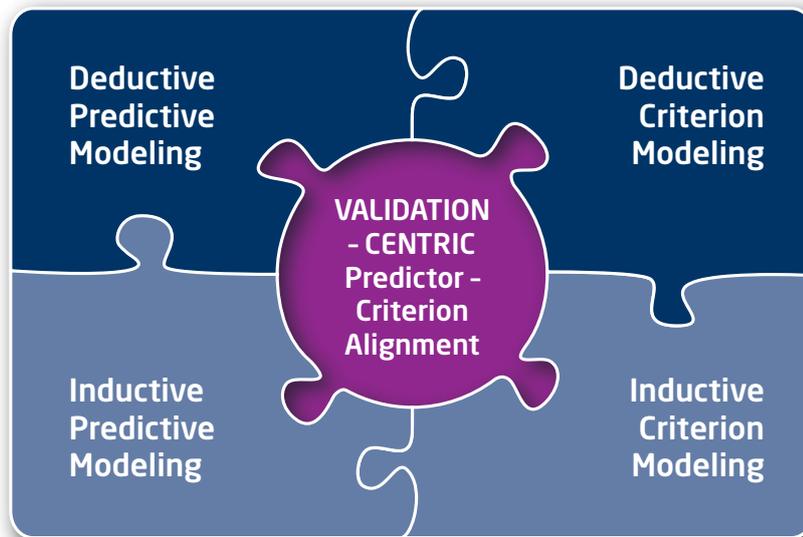
Saville Consulting also focuses on creating assessment tools which are easy to use, acceptable to participants, attractive, and highly applicable in today's modern workplace. Thus, the administration, scoring, and reports for Wave Performance 360 were designed to ensure that its validity is easily accessible and understood by the user.

By achieving these goals, the Wave Performance 360 questionnaire is aimed at providing real value to organizations through assessing characteristics directly related to individuals' performance and potential.

Introducing Saville Consulting Wave Integrated Development

Saville Consulting Wave assessments were based on both inductive and deductive modeling at both the predictor and criterion ends of assessment.

Figure 10.1 Conceptual Overview of Wave Integrated Development



This is an important part of the philosophy behind the Wave suite of assessment tools: they were designed to have one model of effectiveness at their center and have different tools to measure and forecast this one effectiveness model.

On the one hand, we have what we are trying to forecast at the criterion end, which led to the Wave Performance Culture Framework. The Wave Performance culture framework is directly assessed by a number of assessment measures including Saville Consulting Wave Performance 360 and Job Profiler as well as the Wave Performance Culture Framework Card Deck. On the predictor side, a number of assessments and reports have been developed to forecast how effective an individual is likely to be: Saville Consulting Wave Professional and Focus Styles; Saville Consulting Wave Strengths assessments; and the Saville Consulting suite of aptitude assessments. The alignment of these predictors (e.g., Wave Professional Styles) and the criterion measures (e.g., Wave Performance 360) is important in creating clear high validity in tools which forecast effectiveness and ensures they are effective for the selection and development of people at work.

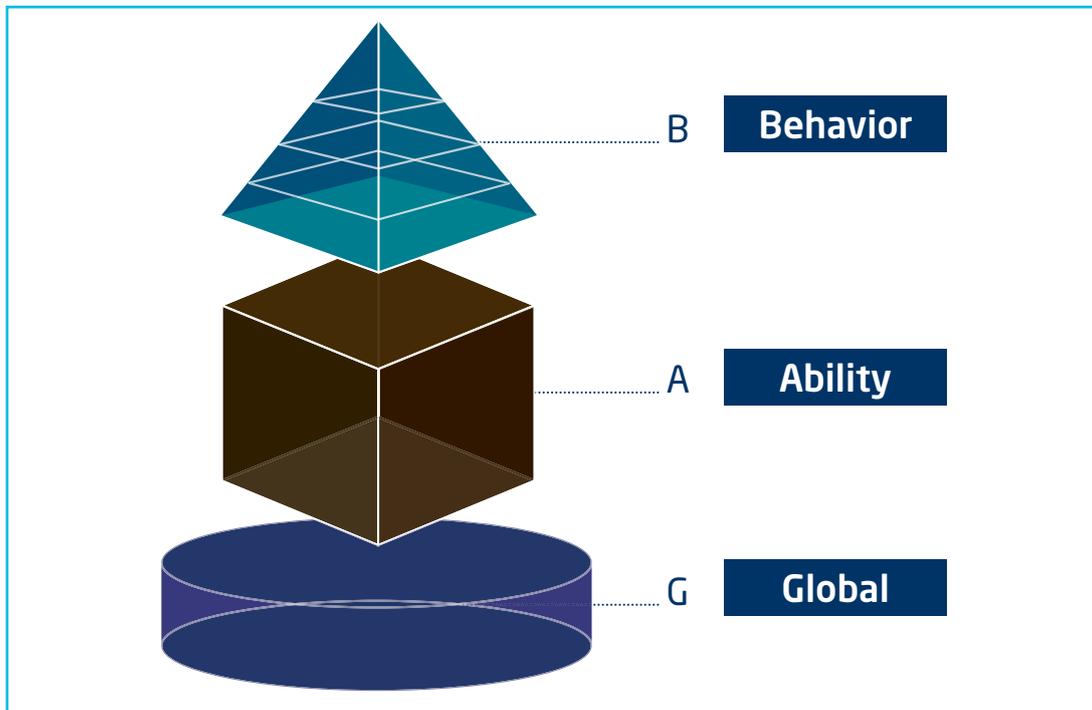
This one effectiveness model seeks to provide a powerful framework for integrated assessment by drawing together a variety of the most important assessment perspectives from high validity measures. The Saville Consulting Performance Culture Framework is a combined model of effectiveness which was designed to cover behavioral competencies, personality traits, cognitive abilities as well as overall (or global) aspects of effectiveness at work.

The criterion side forms the basis of the Performance Culture Framework and is assessed directly using the card deck in addition to Performance 360 and Job Profiler online assessments. It assesses three components of effectiveness: Behavioral components

which cover behavioral strengths or competencies; Ability components covering different areas of intellect that people are more or less effective at applying in their work; and Global components which assess overall aspects of how effective individuals are generally in their work (refer to Figure 10.2).

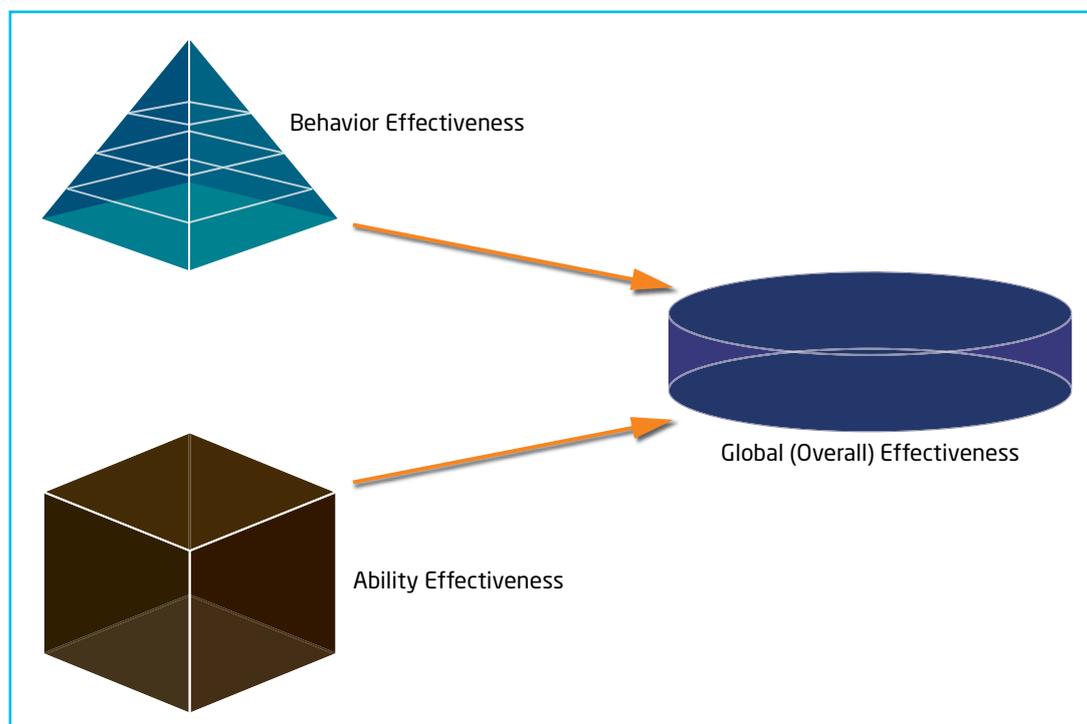
From the perspective of multi-stakeholder 360, the elaboration of the model to include ability and global components is important in giving a more complete picture of effectiveness at work. The inclusion of the ability components, for example, provides rater perspectives on how effective an individual is at 'Working with Words' or 'Working with Numbers'. While an individual may have a high score on a verbal aptitude test such as Saville Consulting Verbal Analysis Aptitude, the 'Working with Words' rating provides additional information on whether this is seen as something an individual is actually effective at. Similarly, the global area can help to provide an individual with an understanding of how effective they are seen to be overall by different raters and rater groups.

Figure 10.2 An Overview of the Structure of the Wave Performance Culture Framework



In Performance 360, behaviors and abilities which were effective in underpinning overall effectiveness at work were modeled and selected. This underpinning of global effectiveness is a fundamentally important part of the assessment model (refer to Figure 10.3). If ability or behavior components assessed do not correlate with overall or global effectiveness measures at work, then they are likely to be less relevant and useful for the selection, development and talent management of people at work.

Figure 10.3 Creating Behavioral and Ability Measures which underpin Global Effectiveness



Thus the predictor measures were designed in the Wave Styles, Strengths and Saville Consulting Aptitude assessment suites to correlate with specific aspects of the Performance Culture Framework as measured in Performance 360. Performance 360 then is a tool which has a model of how its components should forecast each other (ability and behavior correlating with global performance) and a detailed model of how predictor measures will forecast each and every area assessed in the questionnaire.

10.2 Important Perspectives Related to Wave Aligned Model Development

The development of the Performance Culture Framework B-A-G model incorporated a wide range of contemporary theories and models. Some of the more important inputs are described below. Whether the input originates primarily at the criterion end or the predictor end is indicated in brackets.

The Five Factor Model (FFM) or the 'Big Five' Personality Factors (Predictor)

The broad factors known as Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism form the FFM. Barrick & Mount (1991) outlined the research base for these broad trait factors, and traced back their origins to the work of Norman (1963). Over the years different names have been used for what is now understood to be essentially the same construct set. Neuroticism, in the wake of Positive Psychology, is increasingly referred to as Emotional Stability or Confidence, while Agreeableness

and Openness to Experience are sometimes measured through their opposite pole, e.g., Independence and Conventionality, respectively. The NEO-PI-R questionnaire (Costa & McCrae, 1992) is perhaps the most commonly used FFM centric questionnaire.

Alpha and Beta (Predictor)

Digman (1997) investigated the higher-order factor structure of the FFM and found that Agreeableness and Conscientiousness (usually together with Emotional Stability) formed a higher-order factor he called 'Alpha' while Extraversion and Openness to Experience formed a 'Beta' factor. The sections on the left of the Wave Wheel (Figure 10.4) from Structured to Evaluative broadly cover the Alpha supra-factor, and the sections on the right from Investigative to Driven broadly cover the Beta supra-factor.

Hierarchical Model of Abilities (Predictor)

Vernon (1950) proposed a hierarchical model of cognitive abilities that linked the Spearman (1904) model of a general intellectual factor 'g' with the Primary Mental Abilities of Thurstone (1936). The Saville Consulting aptitude assessments are designed to measure verbal, numerical and checking abilities related to Vernon's Verbal-Educational factor while diagrammatic/abstract, spatial and mechanical abilities are related to the Practical-Technical factor in Vernon's hierarchical model. The hierarchical nature of Vernon's model highlights that while we can have differential intelligences, with many individuals being better at some abilities than others, abilities also tend to correlate with each other and have a structure. This allows for the creation of reliable and valid overall reasoning scores when several abilities are jointly assessed.

The Swift series of tests measure higher-order abilities for particular groups of job roles with good reliability in using an appropriate combination of abilities to measure an overall score, while single tests measure individual ability areas. In Performance 360, the structure of abilities is replicated at the criterion end, with the model being split into Working with Information (aligned toward Vernon's Verbal-Educational factor) and Working with Things (aligned to Practical-Technical of Vernon). These are broken down further into Working with Words, Numbers and Details under Working with Information, and Working with Systems, Shapes and Equipment under Working with Things. These directly align with the Saville Consulting aptitude tests that have been created to forecast these effectiveness areas (Working with Words is underpinned by Verbal aptitude tests, Working with Numbers is underpinned by Numerical aptitude, Working with Details is underpinned by Error Checking tests, Working with Systems by Diagrammatic and Abstract tests, Working with Shapes with Spatial tests and finally Working with Equipment is underpinned by Mechanical aptitude assessments).

The Ability component of the B-A-G model is important for job requirements profiling and aptitude test validation work. However, in 360 feedback processes the Abilities section is often omitted as these are less relevant for development in many managerial and professional roles with Documenting Facts, Interpreting Data and Checking Details representing behavioral counterparts to Working with Words, Numbers and Details. Nevertheless, for some job roles the assessment of the ability areas are a fundamental and useful component of 360 feedback and the provision of the ability areas is an important feature of Performance 360 as well as the BAG framework.

The Big One (Predictor)

Since the original development of Wave, Musek (2007) has published a paper which aligns to Wave's construction approach in identifying a 'Big One' of personality which is superordinate to Alpha and Beta. This is an extremely broad overall factor of personality, and it would be unusual to find someone to be high on every aspect of it. Some people will be high on both Alpha and Beta, or at the next subordinate level (e.g., FFM), some people profile as Extrovert, Agreeable, Emotionally Stable, Conscientious and Open to Experience.

Overall Performance Measurement and the Structure of Criteria (Criterion)

The hierarchical structure of the criterion space of work performance was an essential consideration in the development of the Performance Culture Framework. Work criteria tend to correlate with each other, that is, people rated high on one criterion tend to be rated high on other criteria.

There have been concerns about what an overall, large first factor that accounts for a large proportion of variance in the criteria might be measuring. A particular concern is that it may represent construct irrelevant variance and be composed of a number of biases which cause the criteria to interrelate, which do not genuinely reflect the performance of individuals across the criteria. Such construct irrelevant variance may reflect, for example, halo and other biases. However, when halo and certain biases are removed the variance in the first factor largely remains (Viswesvaran, Ones, & Schmidt, 1996), which supports the factor's construct validity in measuring a general factor of performance at work.

There are also correlations between this overall effectiveness rating factor and other independent measures of performance. Initially, two items were developed measuring Proficiency and Promotability inspired by the work of Nyfield et al. (1995). For Project Epsom a three item scale was created covering Accomplishing Objectives, the heart of performance, alongside Applying Specialist Expertise and Demonstrating Potential. A nine-item scale that gathers responses for each definitional component separately was also developed for research purposes.

There is more separation and less positive correlation within predictor measures (especially where predictors have an ipsative score component) than within criteria; as a result we can expect generally to extract fewer (and therefore differently composed) factors from the criterion end than from the predictor end given the same rules for determining the number of factors. Perhaps, this is one reason why prior to the construction of the Wave suite the alignment between the criterion space and predictors has been relatively limited.

The 'Great Eight' Competencies (Predictor and Criterion)

Kurz & Bartram (2002) expanded the FFM by adding three new categories (Analysis, Need for Power and Need for Achievement) into the Great Eight competencies. The Great Eight are designed to account for individual differences in work performance rather than measure personality traits per se. The authors defined competencies in relation to their significance for performance at work as "sets of behaviors that are instrumental in the achievement of desired results or outcomes".

This definition of competencies has important implications for the development of the Wave Performance model as it indicates that there should be a positive association between competencies and achievement of desired results and outcomes. For this to be the case, competencies should correlate with overall measures of performance.

The origins of the model can be traced back to the meta-analysis work of Robertson & Kinder (1993) who created a criterion taxonomy to classify various 'local' performance criteria. This led to the development of a range of competency inventories enabling large scale validation work like the international validation study by Nyfield et al. (1995).

Amendments were made to the Saville Consulting Wave criterion model and predictor facets based on an understanding of how the different Great Eight constructs relate to, and fail to relate to, overall performance positively (see Bartram 2005 for example). In particular, Supporting and Cooperating and Adapting and Coping were amended to focus on behavioral aspects that were hypothesized to better underpin overall measures of performance.

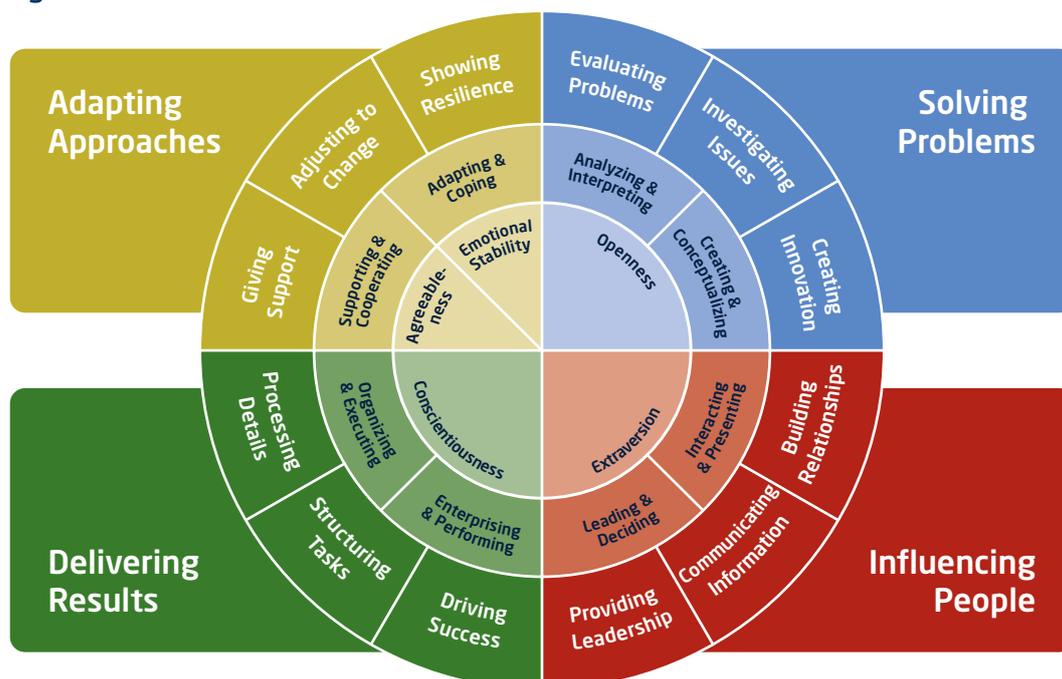
The Four Clusters of Performance and Potential (Predictor and Criterion)

The clusters bring together the diverse range of theories and assessment approaches from the predictor and criterion domain into one model which is designed to underpin performance and potential at work.

The four clusters of Saville Consulting Wave serve as a generic model of work performance across all levels of work complexity and job level.

In practice, the criterion-related validity evidence between the predictors and criteria was the initial basis of the alignment to create the overall Saville Consulting Performance Culture Framework.

Figure 10.4 The Wave Wheel



Twelve Behavior Sections of Performance and Potential (Predictor and Criterion)

The Wave model is hierarchical and the four behavioral clusters are each broken down into three sections. The sections in Saville Consulting Wave are transparently mapped to the Big Five and Great Eight constructs yet are designed to provide wider and more detailed coverage that better reflects the complexity of people and jobs. In Figure 10.4, on the outside, there are the four wave clusters of effectiveness. Moving into the outside of the circle itself, we find the 12 Wave sections. The next circle in towards the center illustrates how the Wave sections and clusters relate to the Great Eight. Finally, the center of the circle shows the FFM, illustrating how the model relates to the Big Five personality traits.

10.3 Construction Phases of Wave Performance 360

The Wave Performance 360 questionnaire was constructed as part of the development process of the Wave assessment suite. Figure 10.5 shows the key initial steps in the construction and standardization of Wave Professional Styles and Performance 360. The steps which specifically relate to the development of Wave Performance 360 are shown in blue in the figure and are explained in this section.

Initial Wave Criterion (Competency) Model

Based on the literature review of common behavioral criteria and popular competency models as well as the latest research in normal personality theory, and the validity of these different measures, a deductive approach was used to create the Wave Behavioral criterion model. The goal was to use the model to create items that would serve to gather performance ratings for validation purposes. This would enable the use of the validation centric approach to select the most valid predictor items and to ensure alignment and validity between the predictor and criterion models.

The criterion model was written in parallel to the predictor model by four psychologists to produce 44 behavioral competency dimensions. These were aligned to the Wave Professional Styles item bank and grouped into a hierarchical structure that included four clusters and 12 sections.

The criterion rating items had a two-word title in bold followed by a list of three sub-components. The 132 competency facet subcomponents in the initial trial were aligned to the deductive Wave predictor model. Table 10.1 shows the original criterion model with the clusters and sections.

Figure 10.5 Overview of Key Steps in Wave Professional Styles and Performance 360 Development

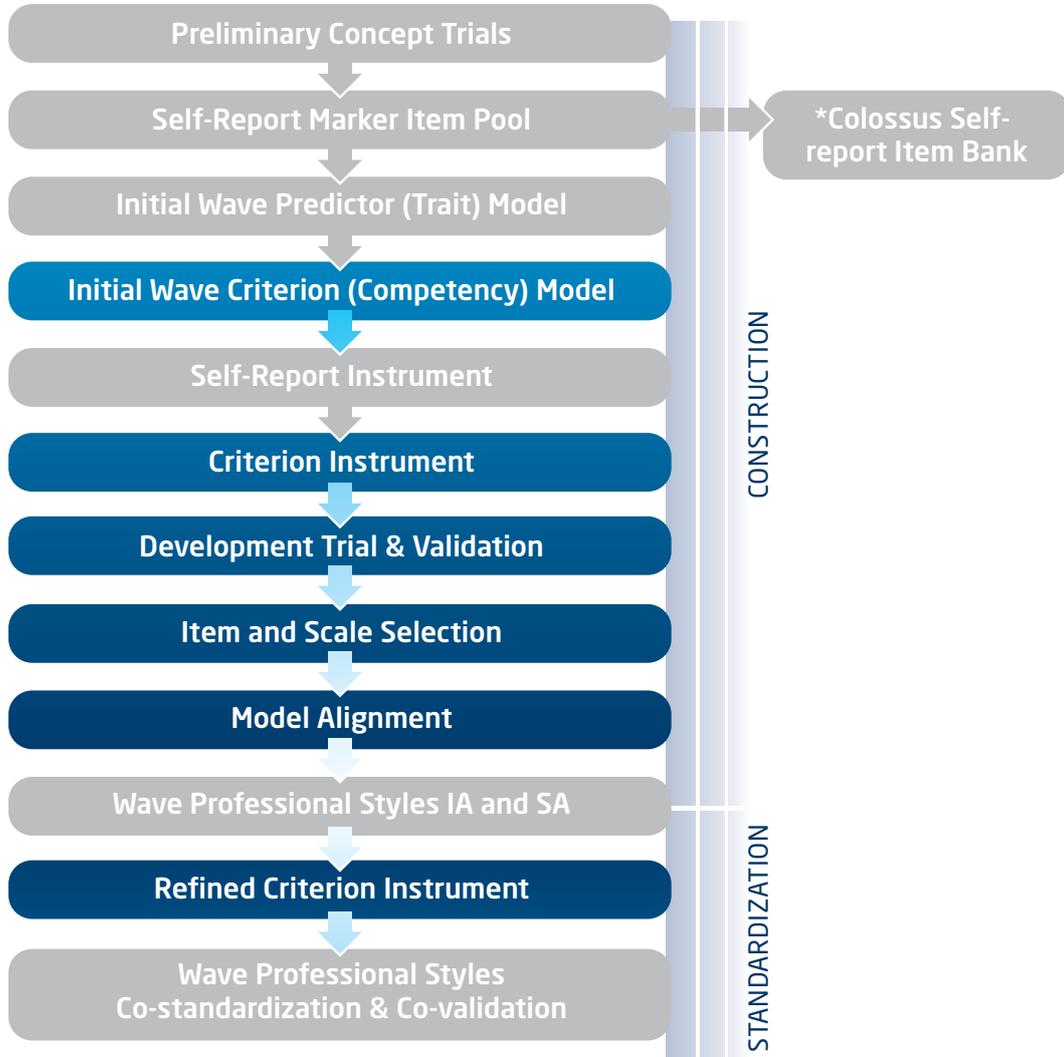


Table 10.1 The Initial Deductive Wave Criterion Model - Included Competency Dimensions before development trial

Competency Dimension	Cluster	Section *
Generating Ideas - e.g., producing ideas; making inventions; adopting radical solutions	Solving Problems	Creating Innovation
Exploring Possibilities - e.g., identifying underlying principles; applying theories; developing concepts	Solving Problems	Creating Innovation
Developing Strategies - e.g., anticipating trends; forming strategies; vision for the future	Solving Problems	Creating Innovation
Providing Insights - e.g., resolving problems; using intuition; making incisive judgments	Solving Problems	Judging Situations
Implementing Practical Solutions - e.g., applying practical skills; utilizing common sense; testing ideas	Solving Problems	Judging Situations
Developing Expertise - e.g., taking up learning opportunities; acquiring knowledge and skills; exploring specialist issues	Solving Problems	Judging Situations
Documenting Facts - e.g., finding facts; probing arguments; writing fluently	Solving Problems	Evaluating Information
Analyzing Situations - e.g., anticipating problems; analyzing information; producing solutions	Solving Problems	Evaluating Information
Interpreting Data - e.g., quantifying issues; evaluating data objectively; utilizing technology	Solving Problems	Evaluating Information
Making Decisions - e.g., assuming responsibility; deciding course of action; standing by decisions	Influencing People	Providing Direction
Leading People - e.g., coordinating groups; providing leadership; controlling things	Influencing People	Providing Direction
Providing Inspiration - e.g., motivating individuals; inspiring people; giving encouragement	Influencing People	Providing Direction
Convincing People - e.g., persuading others; negotiating; shaping opinions	Influencing People	Asserting Views
Challenging Ideas - e.g., questioning assumptions; arguing own perspective; willing to disagree	Influencing People	Asserting Views
Articulating Information - e.g., giving presentations; explaining things; projecting social confidence	Influencing People	Asserting Views
Impressing People - e.g., attracting attention; promoting personal achievement; getting recognition	Influencing People	Communicating with People
Developing Relationships - e.g., making contact; strengthening relationships; networking	Influencing People	Communicating with People
Establishing Rapport - e.g., welcoming people; putting people at ease; making friends	Influencing People	Communicating with People

Competency Dimension	Cluster	Section *
Team Working - e.g., joining team activities; encouraging team contributions; involving others in decisions	Adapting Approaches	Helping People
Understanding People - e.g., listening to people; appreciating others' feelings; understanding motivation	Adapting Approaches	Helping People
Valuing Individuals - e.g., tolerating individual differences; trusting people; showing consideration	Adapting Approaches	Helping People
Resolving Conflict - e.g., calming upset people; handling angry individuals; resolving arguments	Adapting Approaches	Coping with Stress
Conveying Self-Confidence - e.g., valuing own contributions; projecting inner confidence; determining own future	Adapting Approaches	Coping with Stress
Coping with Pressure - e.g., tolerating stress; coping with important events; appearing calm under pressure	Adapting Approaches	Coping with Stress
Thinking Positively - e.g., projecting cheerfulness; being optimistic; recovering from setbacks	Adapting Approaches	Adapting to Change
Inviting Feedback - e.g., encouraging critical thinking; encouraging feedback; acknowledging criticism	Adapting Approaches	Adapting to Change
Embracing Change - e.g., tolerating uncertainty; coping with change; adapting to new circumstances	Adapting Approaches	Adapting to Change
Checking Details - e.g., identifying errors; ensuring accuracy; ensuring high quality	Delivering Results	Delivering Results
Meeting Timescales - e.g., keeping to schedule; meeting deadlines; finishing tasks	Delivering Results	Delivering Results
Following Procedures - e.g., adhering to rules; following instructions; minimizing risks	Delivering Results	Delivering Results
Organizing Resources - e.g., identifying requirements; planning activities; managing projects	Delivering Results	Structuring Tasks
Upholding Standards - e.g., acting with integrity; maintaining confidentiality; behaving ethically	Delivering Results	Structuring Tasks
Completing Tasks - e.g., working efficiently; utilizing time effectively; multi-tasking	Delivering Results	Structuring Tasks
Taking Action - e.g., investing energy; using initiative; sustaining physical activity	Delivering Results	Striving for Success
Pursuing Goals - e.g., acting with determination; persisting through difficulties; achieving results	Delivering Results	Striving for Success
Tackling Challenges - e.g., identifying business opportunities; outperforming competitors; generating profit	Delivering Results	Striving for Success

*Section names used here may be unfamiliar to current users as they were draft names used at this stage in the Development of Wave Professional Styles and were later finalized.

Following on from the work of Nyfield et al. (1995) who differentiated job proficiency from promotability, two aspects of job proficiency and potential for progression formed the initial basis of the criterion assessment that later developed into the Global Cluster of the Wave Performance 360 questionnaire.

Criterion Instrument

A work effectiveness inventory was created using the behavioral items in Figure 10.1 together with the two items of job proficiency and potential for progression. The rating scale was a seven-point Likert-type scale ranging from Extremely Ineffective (1) to Extremely Effective (7). There was also a 'no evidence' option for each rating.

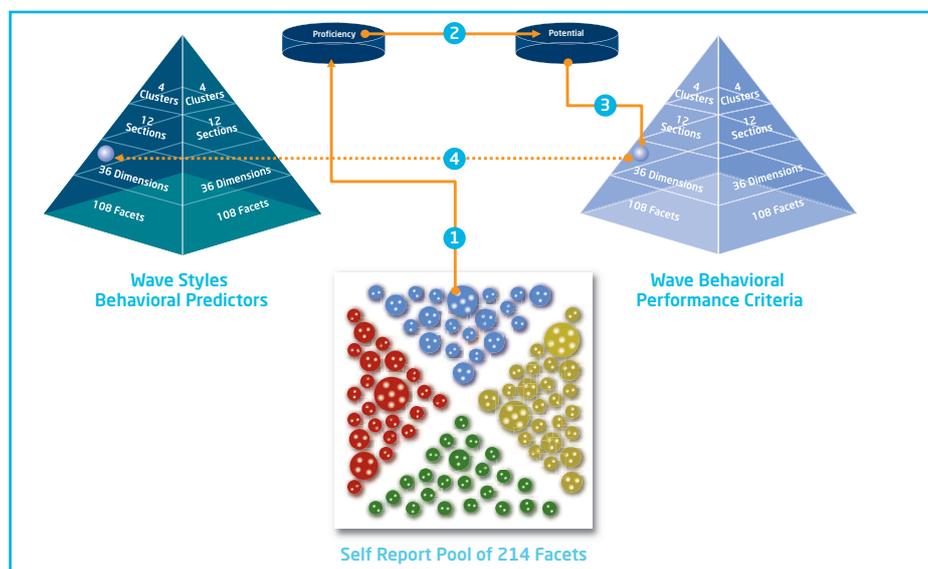
Development Trial & Validation

The work criterion instrument was administered online to a manager, colleague, friend or partner of each Wave Professional Styles questionnaire respondent in the initial trial. 1,011 respondents from over 50 countries completed the Wave Styles Development Trial Self-Report Questionnaire. Between 386 and 394 of these respondents also had external ratings of performance from the work effectiveness inventory (the criterion instrument). This number varied due to the 'no evidence' option on the criterion rating instrument.

Item and Scale Selection

The primary objective of this stage was to select the most valid items and scales from within the predictors and criteria. This process aligned the Wave predictors and criteria and thereby created a model that would maximize criterion-related validity. Each criterion dimension was reviewed in turn to examine how the a priori matched facets correlated with ratings of overall proficiency and potential. The correlation between the facet and its matched criterion was also examined. Facets that correlated best with the overall proficiency and potential and with their matched criterion were selected for inclusion. Refer to Figure 10.6 which provides a simplified overview of the item selection process.

Figure 10.6 Model Creation



1. For each criterion dimension, the facets in the self-report (predictor) model that were aligned to it were examined. For instance, for the criterion dimension of Generating Ideas, the matched self-report facet/items in the area of Creativity and Innovation were examined. The first step was to check that the facets/items correlated with the raters' external view of the individual's overall job proficiency.
2. Next, the facet was considered for selection if it correlated strongly with the future progression criteria.
3. Then the facet's correlation with its matched criterion was examined, e.g., Creative items with Generating Ideas competency.
4. After passing these steps, the most valid facets were placed into the Wave predictor model and selected for the Wave Professional Styles questionnaire. For the Generating Ideas criterion this first facet was called Creative.
5. The process continued by selecting the next most valid facet. For the Generating Ideas competency, this second facet selected was called Original. The next most valid facet for the Generating Ideas competency was called Radical. For each Wave competency, the three facets that correlated highest with overall performance and proficiency as well as its aligned competency were selected to form the Wave Styles dimension.

This process continued until all the competencies were examined and all the Wave Styles (trait) dimensions created. In practice, eight of forty-four competency criteria were dropped based on their overlap with other competency dimensions and their correlations with overall performance and potential. This led to an aligned predictor and competency model with four clusters, 12 sections, 36 dimensions and 108 facets. For more details on the development of the matched predictor and competency model, refer to the Wave Professional Styles Handbook. Table 10.2 shows the top three levels of the final competency model which the Wave Performance 360 questionnaire is based on.

In the process of alignment, eight dimensions were deleted from the model. These dimensions are given in Table 10.3.

Table 10.2 The Final Criterion Model - the top three levels of Wave Performance 360

Criterion Cluster	Criterion Section	Criterion Dimension
Solving Problems	Evaluating Problems	Examining Information Documenting Facts Interpreting Data
	Investigating Issues	Developing Expertise Adopting Practical Approaches Providing Insights
	Creating Innovation	Generating Ideas Exploring Possibilities Developing Strategies
Influencing People	Building Relationships	Interacting with People Establishing Rapport Impressing People
	Communicating Information	Convincing People Articulating Information Challenging Ideas
	Providing Leadership	Making Decisions Directing People Empowering Individuals
Adapting Approaches	Showing Resilience	Conveying Self-Confidence Showing Composure Resolving Conflict
	Adjusting to Change	Thinking Positively Embracing Change Inviting Feedback
	Giving Support	Understanding People Team Working Valuing Individuals
Delivering Results	Processing Details	Meeting Timescales Checking Things Following Procedures
	Structuring Tasks	Managing Tasks Upholding Standards Producing Output
	Driving Success	Taking Action Seizing Opportunities Pursuing Goals

Table 10.3 The Deductive Wave Criterion Model - Deleted Competency Dimensions

Competency Dimension	Cluster	Section *
Investing Trust - e.g., believing in people; giving autonomy; trusting others' intentions	Adapting Approaches	Facilitating Work
Conveying Enthusiasm - e.g., being positive; generating excitement; creating enthusiasm	Influencing People	Building Relationships
Accepting Responsibility - e.g., acting with integrity; admitting mistakes; honoring commitments	Delivering Results	Executing Assignments
Fulfilling Obligations - e.g., being reliable; conforming to organizational values; showing loyalty	Delivering Results	Executing Assignments
Outperforming Competitors - e.g., being competitive; grasping opportunities; winning business	Delivering Results	Delivering Results
Developing Others - e.g., giving encouragement; motivating people; enabling development	Influencing People	Asserting Views
Pioneering New Methods - e.g., challenging convention; improving work methods; designing new approaches	Solving Problems	Developing Concepts
Developing Self-Insight - e.g., developing self-awareness; reflecting on experience; learning from experience	Solving Problems	Developing Concepts

**Names used here may be unfamiliar to current users as they were draft names used at this stage in the Development of Wave Professional Styles and were later finalized.*

Refined Criterion Instrument

The criterion instrument was refined to cover the Behavior areas and Global (Overall) Effectiveness areas. In addition, the Ability areas were added to create the full Wave Performance Culture Framework (Figure 10.5).

The original work effectiveness survey was redesigned to collect criterion ratings only on the 36 behavioral work effectiveness areas selected following the standardization trial.

The Ability Profile was added at this stage to measure effectiveness areas that mirror the Saville Consulting Aptitude portfolio. Table 10.4 shows the six Ability items constructed and their corresponding aptitude areas. These aptitude questions aim to provide additional information on how an individual views themselves and how their colleagues (boss, peer and report) see them in terms of different ability areas. Individuals can perform well on a specific aptitude test, however, it is a separate (if related) question whether an individual who scores highly on an aptitude test can realize their potential and perform effectively by applying their intellect at work. This profile was designed to be an optional section in the Wave Performance 360 questionnaire. It can be omitted when certain aptitude areas are not relevant to the job role or when a multi-rater feedback on aptitude areas is not required.

Table 10.4 Items in the Wave Performance 360 Ability Profile

Wave Performance 360 Ability Items	Aptitude Area
Working with Words - e.g., Understanding Word Meaning; Comprehending Text; Making Verbal Inferences; Evaluating Written Materials; Comparing Arguments	Verbal
Working with Numbers - e.g., Understanding Tables; Comprehending Graphs; Making Numerical Inferences; Evaluating Quantities; Comparing Data	Numerical
Working with Details - e.g., Checking Letters and Text; Checking Numbers and Tables; Checking Codes and Symbols; Identifying Mistakes; Classifying Information	Error Checking
Working with Systems - e.g., Understanding Logical Rules; Comprehending Process Diagrams; Identifying Causes; Finding Faults; Comparing Flowchart Sequences	Diagrammatic
Working with Designs - e.g., Estimating Lengths and Angles; Recognizing Rotated Shapes; Visualizing Three-dimensional Objects; Inspecting Objects; Designing Things	Spatial
Working with Equipment - e.g., Understanding Mechanical Problems; Comprehending Physical Principles; Estimating Movement of Objects; Using Tools; Operating Machinery	Mechanical

Global Areas

Following on from the work of Nyfield et al. (1995) who differentiated job proficiency from promotability, the proficiency dimension was further separated into two aspects, Applying Specialist Expertise and Accomplishing Objectives. Potential for progression was redefined as a more general concept of demonstrating potential. Table 10.5 shows the final Global items in the Wave Performance 360 questionnaire.

Table 10.5 Global Items in the Wave Performance 360

Wave Performance 360 Global Items
Applying Specialist Expertise - e.g., Utilizing Expert Knowledge; Applying Specialist Skills; Sharing Expertise
Accomplishing Objectives - e.g., Achieving Personal Targets; Contributing to Team Objectives; Furthering Organizational Goals
Demonstrating Potential - e.g., Seeking Career Progression; Demonstrating Capabilities Required for High Level Roles; Showing Potential for Promotion

Each Global Section (item) is underpinned by three dimensions. These dimensions are designed to provide a wide span of application to maximize their applicability and impact across an organization. For example, in Accomplishing Objectives, the facets range from focusing on achievement of personal targets, through to contributing towards the realization of team objectives, and finally towards providing an input to organizational goals.

The three sections are intended to be aggregated to one measure of Global (overall) effectiveness. An individual who is effective in Applying Specialist Expertise, Accomplishing Objectives and Demonstrating Potential will be seen as effective at work overall. That said, specific roles may be more focused on one of these specific dimensions and profiling individuals against these areas helps individuals understand their overall strengths.

Refined Rating Scale

The original survey used a seven-point scale ranging from Extremely Ineffective (1) to Extremely Effective (7) with a 'no evidence' option for each rating. Following the standardization trial, it was decided to drop the 'no evidence' option for several reasons. Firstly, this option was found to be rarely used as behavior dimensions in the Wave model were designed to be applicable across a very wide range of roles. Secondly, it was believed that a lack of evidence in these generic behaviors would mean a lack of positive display of these behaviors. Thirdly, 'no evidence' ratings were found to reduce the statistical power in validation studies due to decreased sample size (as cases with a single use of the 'no evidence' option would be excluded completely from studies). Lastly, when the 'no evidence' option was placed at the lower end of the rating scale (i.e., 0) in our analyses, higher validities were obtained than when this option was completely removed from the rating scale. As a result, an 'unsure' option replaced 'no evidence' in the refined instrument. The 'unsure' option was placed in the middle of the seven-point scale indicating that this option should be neutral, rather than leaning towards positive or negative display of behaviors. Therefore, the rationale for the revised Performance 360 rating scale is that on the one hand there is positive evidence of an individual's effectiveness. This positive rating evidenced by a rating of Fairly Effective or higher is stronger than giving an individual a rating of 'unsure'. 'Unsured' of course could be due to a lack of evidence, but a lack of evidence is weaker than positive evidence. Likewise, negative evidence in the form of a rating of Fairly Ineffective or lower is more negative than lacking evidence or being 'unsure' of whether the individual is effective or ineffective.

Dual Scoring and Reporting

Raw ratings are first collected at the Wave Dimension level and then aggregated to form Section scores. Two different formats of scoring and reporting were designed. The first format, direct reporting of scores, is an aggregation of raw ratings by different raters to form group means (boss, self, peer and report). This format of reporting retains the terminology used in the rating scale so as to provide straightforward interpretation of scores that can be easily understood, for example, feedback can be given to an assessee that their peers overall gave them a rating of Very Effective on Examining Information. The second format, benchmark comparison, is to compare the individual's scores against a relevant norm group. For detailed information on these two formats of scores, see Chapter 11 'Scoring and Profiling' in this handbook.

Wave Performance 360 Reports were then built to be straightforward, clear, attractive, and easy to understand. See 'Sample Reports' chapter (Chapter 6) for further information.

Norms

The initial norm created for Wave Performance 360 was based on the Wave Professional Styles standardization sample. As an ongoing development program, two norms are currently available to suit two different levels of management. Further discussion about norms can be found in the 'Norms' chapter (Chapter 12).

11.0 Scoring & Profiling

Two different formats of scores are presented in the Wave® Performance 360 Report.

Direct Reporting

The first format of scores is direct reporting based on the seven effectiveness categories presented to the raters in the Performance 360 online assessment. These range from 'Extremely Effective' to 'Extremely Ineffective'. This reporting format provides a direct report of the rating given by each rater group. It is therefore simple to understand, for example, that the boss has given an assessee a rating of Extremely Effective. This direct form of reporting is straightforward, gives a direct understanding of ratings given and is generally positive, as raters tend to give ratings in the top three rating categories (Fairly to Extremely Effective). The scoring process transforms ratings from the seven categories into numerical values from 1 (Extremely Ineffective) to 7 (Extremely Effective).

Benchmark Comparison

When interpreting the results of Performance 360, it is also important to know how an assessee's score compares with scores obtained by others, i.e., whether it was high, average or low performance in comparison with a relevant group (e.g., Professionals & Managers). In Wave Performance 360, raw scores (scores in the direct reporting format) are benchmarked against an external norm group and shown on a 1 to 10 sten scale. Apart from providing a meaningful benchmark, sten scores also enable comparison with the Professional Styles or Focus Styles Expert Report. More information on norming and the norms available for Performance 360 can be found in Chapter 12 'Wave Performance 360 Norms' of this handbook.

The two formats of scores are shown by rater category on all the profiles except the Behavior Overview Profile, which displays a combined score from all rater categories.

11.1 Behavior Profile and Ability Profile

The Behavior Profile is shown across four pages in the Performance 360 Report. Each page of the profile is devoted to one cluster reporting the scores of the nine Behavior dimensions that sit underneath the cluster.

Figure 11.1 shows the scores an individual received on Examining Information, Documenting Facts and Interpreting Data under the Evaluating Problems section in the Solving Problems cluster.

Direct Reporting

In the example, the individual's Boss rating on the Interpreting Data dimension was Fairly Effective whereas the Self rating was Very Effective. When there is more than one rater in a rater category, ratings are averaged and the averaged score is display on the profile. Where there is a range of ratings, the range is indicated by arrows giving the lowest score within the rater group through to the highest rating. An arrow will only appear left or right

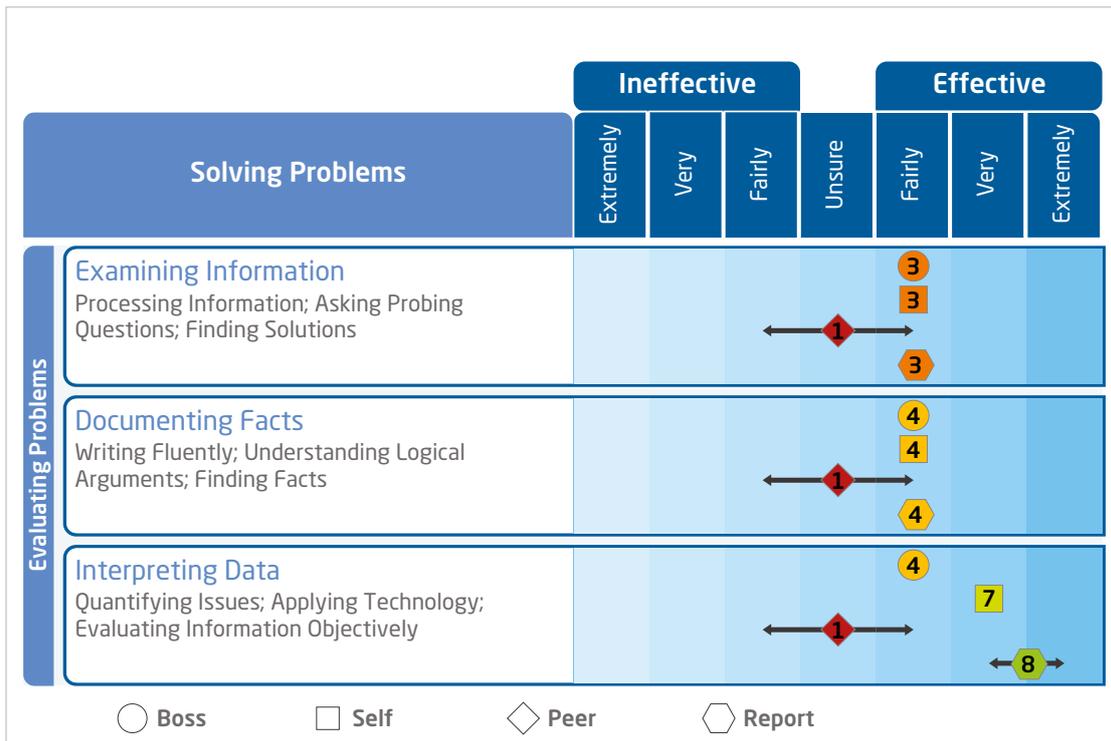
of the score where the difference between the lowest and highest score is at least one score rating (i.e., scores in the direct reporting format rather than sten scores). In the above example, the average of Peer ratings was Unsure, and the lowest rating received was Fairly Ineffective while the highest was Fairly Effective. Similarly, the lowest Report rating received was Very Effective and the highest was Extremely Effective, with the average Report rating falling in between these two points.

Benchmark Comparison

Compared with the norm group, the Boss rating on Interpreting Data was Fairly Low (sten 4) and Self rating was Fairly High (sten 7). The average of Peer ratings was Extremely Low (sten 1) and the average of Report ratings was High (sten 8) in comparison to other individuals.

Where the Wave Performance 360 which includes the Ability Profile is being used, the two formats of scores are displayed in the same way on the Ability Profile. The Ability Profile presents scores on the one ability cluster 'Reasoning at Work' which comprises two sections with three dimensions under each.

Figure 11.1 A sample Behavior Profile showing scores for the Evaluating Problems section under the Solving Problems cluster



11.2 Summary Profile

11.2.1 Global Sections

The first part of the Summary Profile displays scores on the three Global sections: Applying Specialist Expertise, Accomplishing Objectives and Demonstrating Potential.

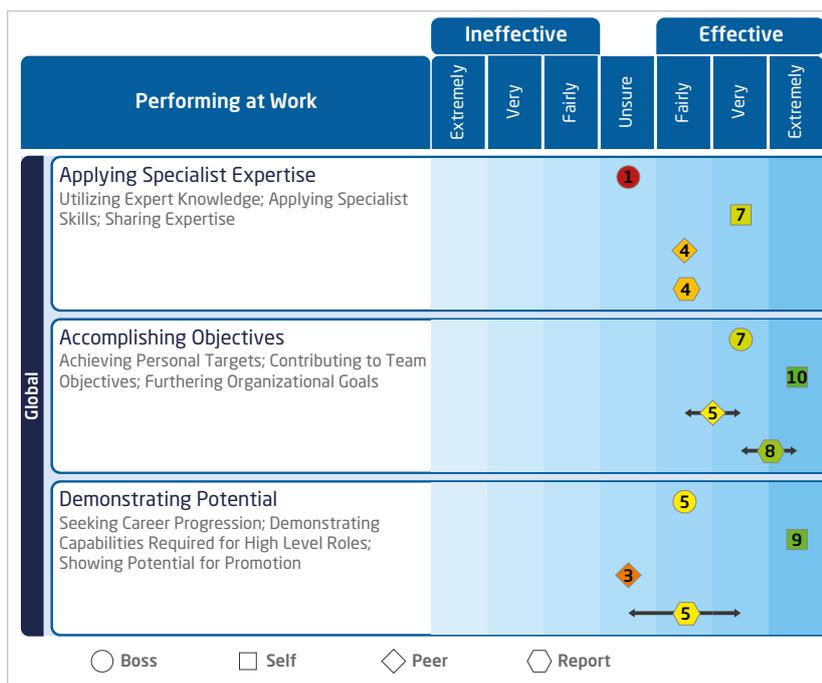
Direct Reporting

In the sample profile below (Figure 11.2), the individual's Boss rating on Accomplishing Objectives was Very Effective and the Self rating was Extremely Effective. The arrows in the Peer and Report ratings indicate that there was a range of ratings in these rater categories. The lowest rating in the Peer ratings was Fairly Effective and the highest rating was Very Effective, whereas the lowest rating the individual received from Report was Very Effective and the highest was Extremely Effective. Because average ratings are calculated by averaging the corresponding numerical values (from 1 to 7) of the rating categories, they can result in non-integers. Therefore, markers can appear in between two rating categories. In the example below, the average rating from Peer lies in between Fairly Effective and Very Effective, and that from Report lies in between Very Effective and Extremely Effective.

Benchmark Comparison

Compared with the norm group, the Boss rating on Accomplishing Objectives was Fairly High (sten 7), Self rating was Extremely High (sten 10), the average of Peer ratings was Average (sten 5) and the average of Report ratings was High (sten 8).

Figure 11.2 A sample Summary Profile showing scores for the three Global sections



11.2.2 Total Ratings

The second part of the Summary Profile presents the total scores for each profile: Behavior, Ability (if used), and Global Ratings. Calculations for each of the total ratings follow the same steps. For the Total Behavior Profile Rating, an average rating across all the 36 behavior dimensions is first calculated for each individual rater (including the assessee). Where there are multiple raters in a rater group, these average ratings are then aggregated to create a total rating score for the rater group. The Total Ability Profile Rating takes the average rating across the six ability dimensions, and the Total Global Rating takes the average of the three global sections.

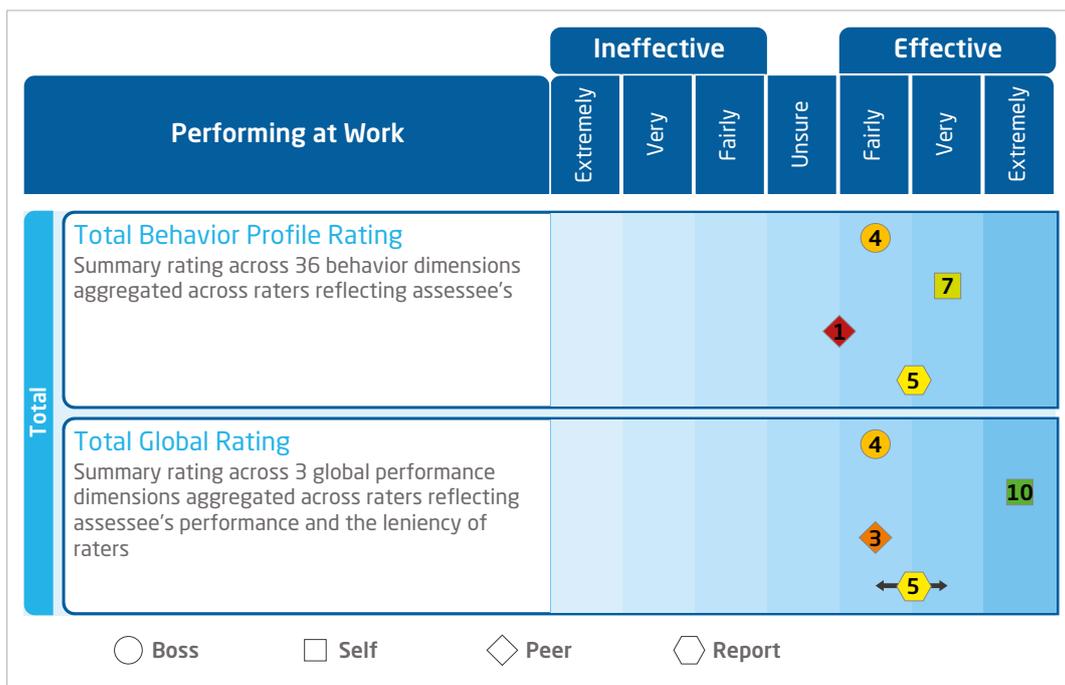
Direct Reporting

In the example below (Figure 11.3), the individual received a Total Global Rating of Fairly Effective from both their Boss and Peer rater groups, and Extremely Effective rating was the average they gave themselves. The Total Global Rating average from the different Reports ranged from Fairly Effective to Very Effective within this rater group with the average total score falling between Fairly and Very Effective.

Benchmark Comparison

Compared with the norm group, the Total Global Rating the individual received from their Boss was Fairly Low (sten 4). The rating the individual gave themselves was Extremely High (sten 10). The total rating from the Peer group was Low (sten 3) and from the Report group was Average (sten 5).

Figure 11.3 A sample Summary Profile showing scores for the Total Ratings



11.3 Behavior Overview Profile

The Behavior Overview Profile is a single-page profile showing the overall scores that an assessee received from all raters on the 12 behavioral sections and 36 behavioral dimensions. The scores from all raters in a rater category are averaged, and then the scores from each rater group are averaged to create an overall score.

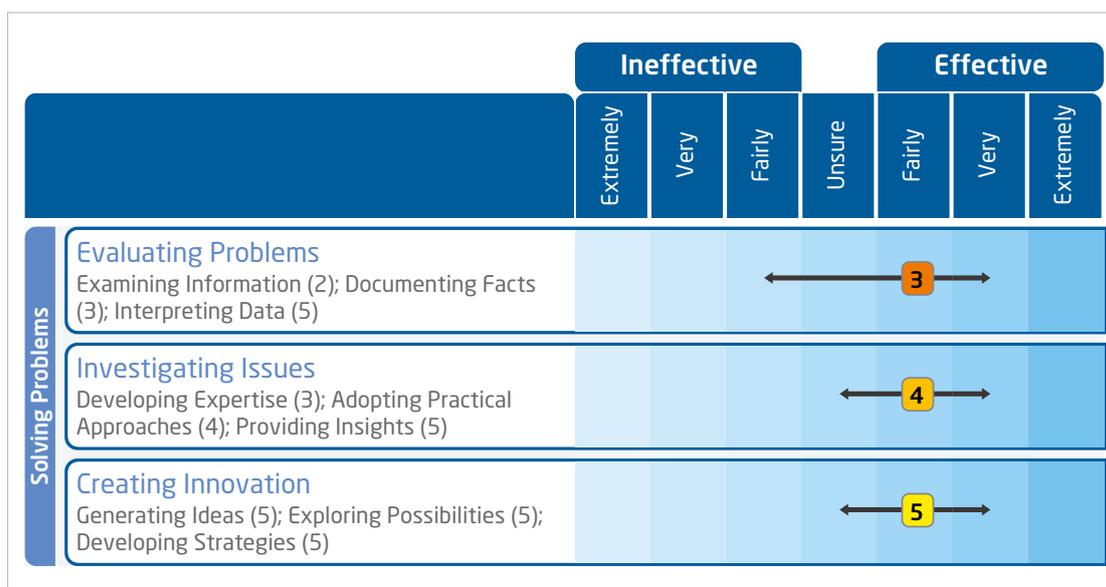
Direct Reporting

In the sample profile shown below (Figure 11.4), the overall score that the assessee received on the Evaluating Problems section was Fairly Effective. The arrows on both sides of the overall score indicate the range of scores given by individual raters. The lowest score the assessee received on Evaluating Problems was Fairly Ineffective and the highest was Very Effective.

Benchmark Comparison

Compared with the norm group, the overall score for Evaluating Problems was Low (sten 3). Scores at dimension level are presented in stens on the left hand side of the profile under their corresponding sections, i.e., these scores are relative to the comparison group. Under the Evaluating Problems section, the Examining Information dimension has an overall score of Very Low (sten 2), Documenting Facts is Low (sten 3) and Interpreting Data is Average (sten 5).

Figure 11.4 A sample Behavior Overview Profile showing overall scores for the Solving Problems cluster



11.4 Rater Comments

All raters, both the assessee and raters, who complete the Performance 360 questionnaire will be presented with three questions at the end where they can provide additional comments. The three standard questions are related to what the assessee could keep doing well at, what they could do less of and what they could improve at.

Comments made by all raters will be displayed on the last three pages of the Performance 360 report. It is important to communicate to all raters in advance that comments will be shown verbatim. If a rater has not made any comments to a question, the text 'No comments were made' will be shown. The Rater Comments taken from a sample report are shown in Figure 11.5. In this example, all raters (Boss, Self, 2 Peers and 2 Reports) who completed the questionnaire have made comments to all the three questions. When there is more than one rater in a rater category, raters' comments are shown in no particular order, i.e., the order is not dependent on the time raters completed the questionnaire or the time they were added to the Performance 360 project. However, each rater will appear as the same individual across the three questions, i.e., Peer 1 is the same Peer rater across all the three questions.

Prior to giving feedback to an assessee, the individual delivering the feedback may want to check that there are no comments from raters which might be considered inappropriate.

Figure 11.5 Sample Rater Comments pages showing the three open text questions

The figure displays three overlapping screenshots of 'Rater Comments' sections. Each section is titled 'Rater Comments' and contains a question followed by responses from different raters.

Top Screenshot: Jo Wilson needs to keep doing well at...

- Boss 1: driving his team to be positive and achieve results, creating a sense of purpose and motivation - bringing in five big projects
- Self 1: Delivering results and growing the revenues
- Peer 1: delivering his res positive and enc
- Peer 2: Seeking out sales Focusing on reve
- Report 1: motivating his te fun
- Report 2: Introducing me to my product know

Middle Screenshot: Jo Wilson needs to do less of...

- Boss 1: Missing key facts or details - keeping client info up to date and keeping consultants in the business up to date with what he and his team are doing in their accounts
- Self 1: Getting distracted by irrelevant detail
- Peer 1: involving experts too late in the process to allow them to input into the design, making up checking their av
- Peer 2: Being political an outweighs the te environment. Car
- Report 1: being negative a openly - falling o reason
- Report 2: Taking over , whe

Bottom Screenshot: Jo Wilson needs to improve at...

- Boss 1: Bit more checking of facts and communication across the business would help and learning to bring in expertise a little earlier on some big projects rather than scope them himself
- Self 1: Winning bigger customers and cross selling within these
- Peer 1: checking things, making sure adminstration is done accurately on client meeting/orders and client enquiries. letting other people know what is going on
- Peer 2: Understanding other people's position and being prepared to be flexible with people and resources. Developing longer term strategies in addition to focusing on the quarterly revenues
- Report 1: checking information and prices on proposals he is signing off- not giving one off deals to clients that are difficult to track
- Report 2: Thinking about how to motivate me and the team. No news is good news! Could be better at anticipating when I actualky do need help/guidance and when I don't

12.0 Wave® Performance 360 Norms

12.1 About Norms

When interpreting the results of an assessment it is often useful to know how each individual's score compares with scores achieved by others. Knowing whether a score is high, low or average compared with others requires that we have a norm group. Norms allow for comparison of an individual's score on an assessment with a relevant comparison group. The use of norms ensures that, when comparing the scores of different individuals, you can be sure you are comparing like with like.

There are various standard scales that could be used to assess individuals in aptitude and behavioral styles assessments. Often, different scales are used for aptitude and behavioral assessments. To allow for a common simple language on both behavioral style and aptitude, sten scores are available. 'Sten' stands for 'standard ten' and stens provide a score which ranges from 1 to 10 with 5 and 6 straddling the average (mean) score. While this provides a simple scale for users, it is also useful to understand how these scores approximate to percentiles in the normal distribution (see Figure 12.1 below).

Figure 12.1 Stens 1 - 10 and their relation to percentiles in the normal distribution

1 - Extremely Low	- performed better than only 1% of comparison group
2 - Very Low	- performed better than only 5% of comparison group
3 - Low	- performed better than only 10% of comparison group
4 - Fairly Low	- performed better than only 25% of comparison group
5 - Average	- performed better than only 40% of comparison group
6 - Average	- performed better than 60% of comparison group
7 - Fairly High	- performed better than 75% of comparison group
8 - High	- performed better than 90% of comparison group
9 - Very High	- performed better than 95% of comparison group
10 - Extremely High	- performed better than 99% of comparison group

For simplicity of use, these are rounded to give whole number percentiles (positive integers), where possible as a multiple of 5 or 10, which are near the center of each sten score. This avoids creating the perception of over accuracy in the score particularly as stens are bands of scores which are subject to a degree of error.

Calculating Sten Scores

When using Wave Performance 360, users do not need to calculate sten scores manually, as our online assessment system (Oasys) does this. However, for those who are interested or would like a reminder, the formulas for calculating sten scores are presented for reference below.

Sten scores are calculated from a person's raw scores on an aptitude or behavioral styles assessment.

To work out a person's sten score, you first need to calculate the Z-score. A Z-score represents how far away a person's score is from the group mean in standard deviation units. The formula to calculate a person's Z-score is as follows:

$$\text{Z-score} = \frac{\text{Individual's raw score} - \text{Mean of the group}}{\text{Standard Deviation}}$$

$$\text{Z-Score} = \frac{x - \bar{x}}{\text{SD}}$$

From this, you can work out a person's sten score. The formula for calculating sten scores is given below:

$$\text{Sten score} = (\text{Z-score} \times 2) + 5.5$$

A sten score gives a rounded representation of a person's score against a benchmark comparison group. One sten score covers half of a standard deviation from the bottom of the score to the top of the sten score.

Standard Error of the Mean (SE_{mean})

Standard Error of the Mean (SE_{mean}) is a measure of how accurate a representation your sample mean is of the 'true' population mean. The larger your sample size, the more accurate it is at representing the true population mean. Table 12.1 demonstrates how SE_{mean} is related to sample size.

There is always a quest within psychometric assessment to have the largest possible numbers for the analysis and interpretation of data. While this is essential for reliability and validity analysis, when considering Standard Error of the Mean, this is not always so necessary. As can be seen in Table 12.1, for a sample size of around 500, the impact of increasing sample size upon Standard Error of the Mean only serves to make an already small error even smaller.

Therefore, although in general we can say that the larger the sample size the better, in terms of normative data collection, having a very large sample is often less important than other considerations. In practice, the most important consideration when collecting normative data is often that the sample should be representative of the population, and this becomes harder to ensure as sample sizes increase.

Table 12.1 Standard Error of the Mean at different sample sizes

Sample Size	SEmean (stems)
50	.29
100	.20
250	.13
500	.09
1000	.06
10,000	.02

Available Norm Groups and their Development

The norm groups currently available for Performance 360 are Senior Managers & Executives (2011) and Professionals & Managers (2011).

In order to calculate these norms, data was drawn from Performance 360 assessments completed on the Oasys platforms. The majority of the sample was drawn from individuals using Performance 360 to assess their development, participants in organizational talent audits, and participants of research projects or Performance 360 training courses. Initial classification into the two groups was done mainly using the assessee’s self-report on level of management responsibility. The Senior Managers & Executives group consisted of assessees who were at one of the following management levels: Board, Enterprise/Corporate, Executive, Director, Vice-President or Senior Manager. The Professionals & Managers group contains assessees who reported themselves as Business Managers, Functional Managers, Group Managers, Managers, Team Leaders, Professional Individual Contributors or at Professional/Specialist level, plus those from the Senior Managers & Executives group.

As the information on level of management responsibility was self-reported, some sense checks were carried out to ensure that both groups contained the targeted assessees. For the Senior Managers & Executives group, assessees who reported themselves less than 27 years of age and with less than 3 years’ total work experience were excluded. Assessees who were less than 24 years of age and with less than a year’s work experience were excluded from the Professionals & Managers group.

The reliability of external ratings was also important. Similar criteria were therefore set for the raters in both norm groups. For the Senior Managers & Executives group, raters who were less than 25 years of age and with less than a year’s work experience were excluded. For the Professionals & Managers group, raters had to be at least 21 years of age and with at least 6 months of work experience in total.

The final Senior Managers & Executives group consists of 778 assessees and 4556 raters, and the Professionals & Managers group contains 2268 assessees and 10324 raters. Full demographic information of the two norm groups is described in the next section.

12.2 Performance 360 Norm Group Description

Norm Group: Senior Managers & Executives (2011)

This sample consisted of 5334 individuals¹ employed in a wide range of job areas/ functions². Approximately 12% worked in Sales & Marketing, 11% in Management, 11% in Education, 10% in Human Resources, 9% in Finance, 7% in Operations, 5% in Law and 5% in Administration. The remaining 30% worked in other areas including Military, Information Technology, Production, Health, Research, Consulting, Engineering, Leisure, Government/ Civil Service, Construction, Transport and Media.

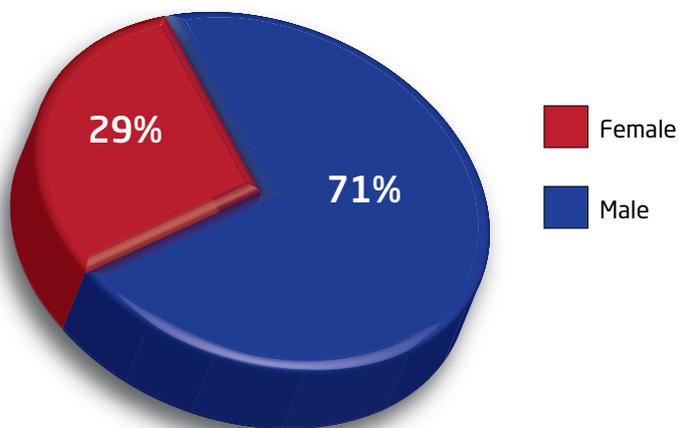
The norm group was based on combination of all raters, including self and other raters. The breakdown for the two rater categories is provided below (with response rates for each biographical section given in the footnotes):

Self-raters

The sample consisted of 778 senior managers and executives who rated themselves in the questionnaire.

Gender³

71% of the sample were male and 29% were female.



Age⁴

The age of the group ranged from 27 to 63 years, with a median age of 43 years.

¹ The number of participants included all raters in a given set of Performance 360 assessment (Bosses, Self, Peers and Reports). Of these, 1564 participants also completed the Performance 360 Ability Profile.

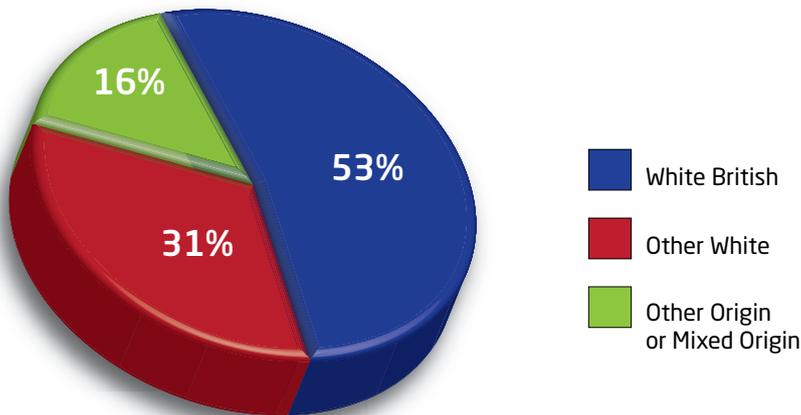
² Based on 69% sample response

³ Based on 100% sample response

⁴ Based on 75% sample response

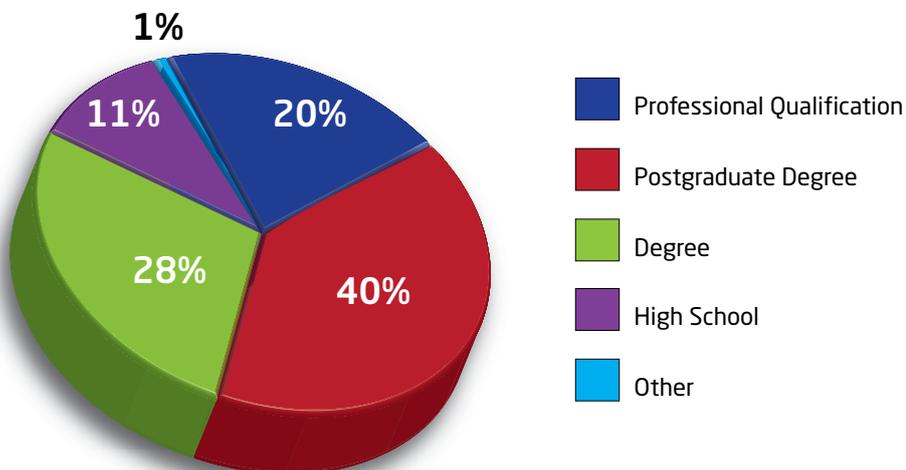
Cultural Background⁵

Approximately 54% of the sample described themselves as UK White and 31% as being from other White backgrounds (including European, American, Australian and New Zealander). The remaining 16% of the sample were of other (e.g., Indian, Hispanic, Black African) or mixed origin (e.g., White and Asian, White and Black).



Education (highest qualification)⁶

Approximately 20% had a professional qualification, 40% had a postgraduate degree, 28% of the group had a degree, 11% were educated to high school, and less than 1% of the group stated 'Not applicable' or 'other'.

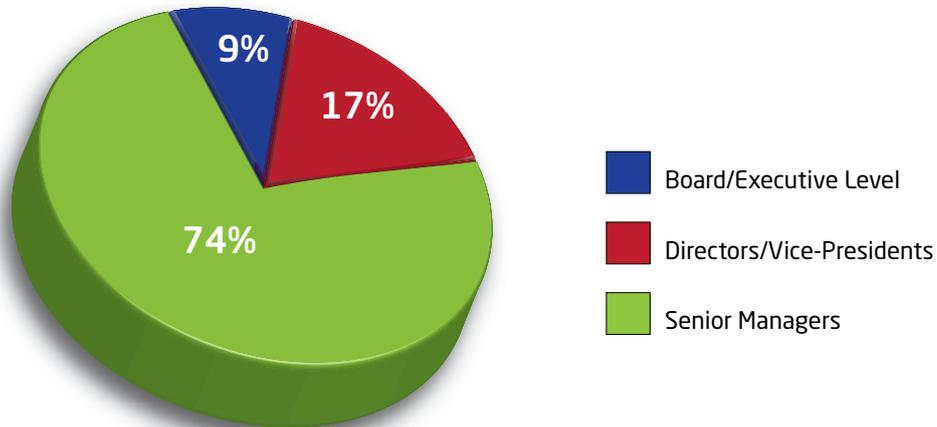


⁵ Based on 59% sample response

⁶ Based on 98% sample response

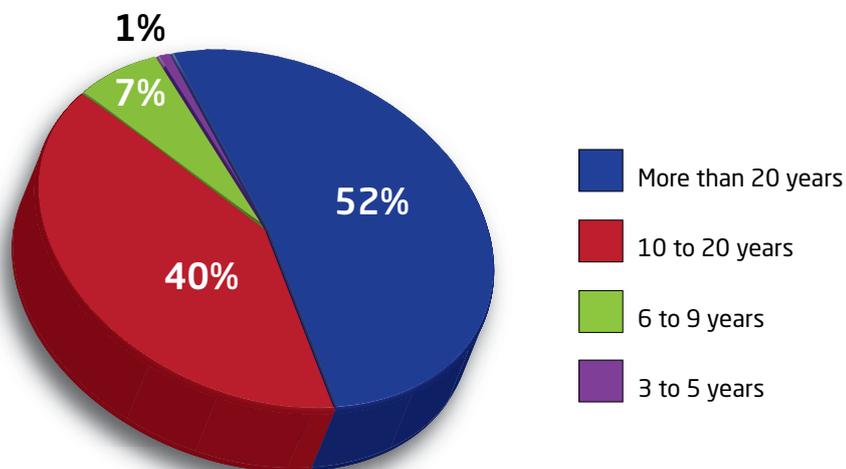
Level of Responsibility⁷

Approximately 9% of the sample were at board or executive level, 17% at director or vice-president level, and the remaining 74% were senior managers.



Work Experience⁸

The overall work experience of the group ranged from 3-5 years (1%) to more than 20 years (52%). 7% had between 6 and 9 years of work experience, and 40% had between 10 and 20 years.



⁷ Based on 100% sample response

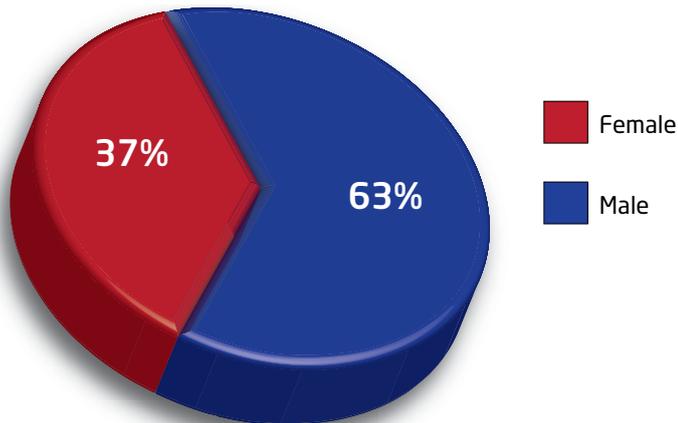
⁸ Based on 100% sample response

Other raters

The sample consisted of 4556 individuals who rated the work performance of a colleague in the questionnaire.

Gender⁹

63% of the sample were male and 37% were female.

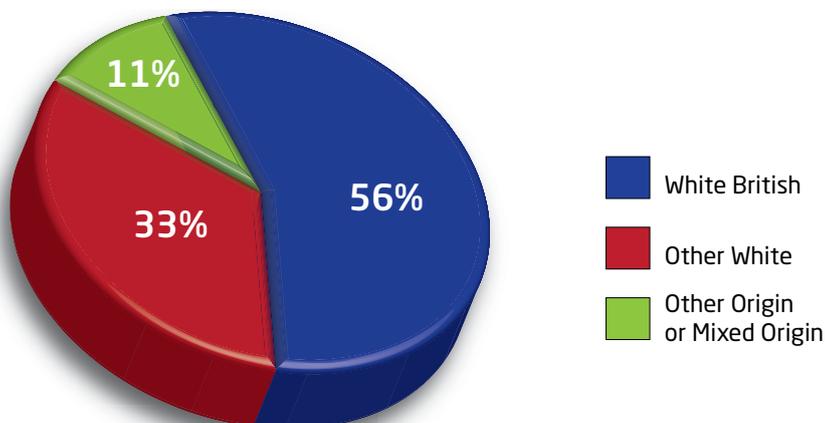


Age¹⁰

The age of the group ranged from 25 to 74 years, with a median age of 44 years.

Cultural Background¹¹

Approximately 56% of the sample described themselves as UK White and 33% as being from other White backgrounds (including European, American, Australian and New Zealander). The remaining 11% of the sample were of other (e.g., Indian, Hispanic, Black African, Black American, Black Caribbean) or mixed origin (e.g., White and Asian, White and Black African, Black African and Asian).



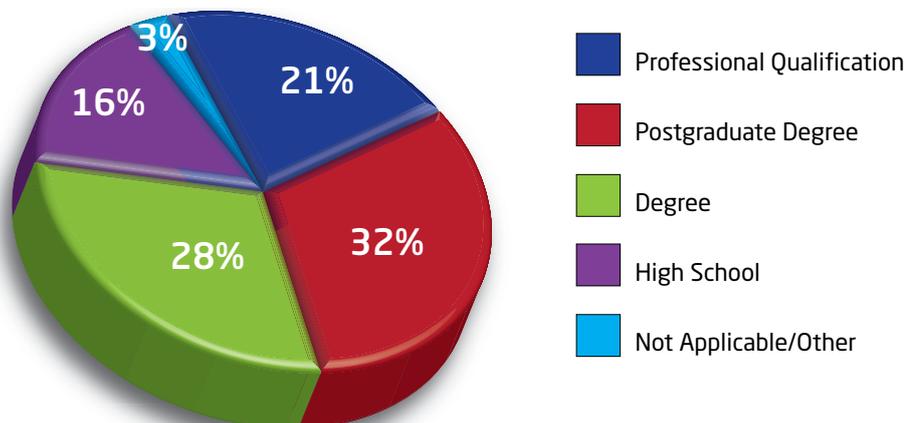
⁹ Based on 69% sample response

¹⁰ Based on 45% sample response

¹¹ Based on 54% sample response

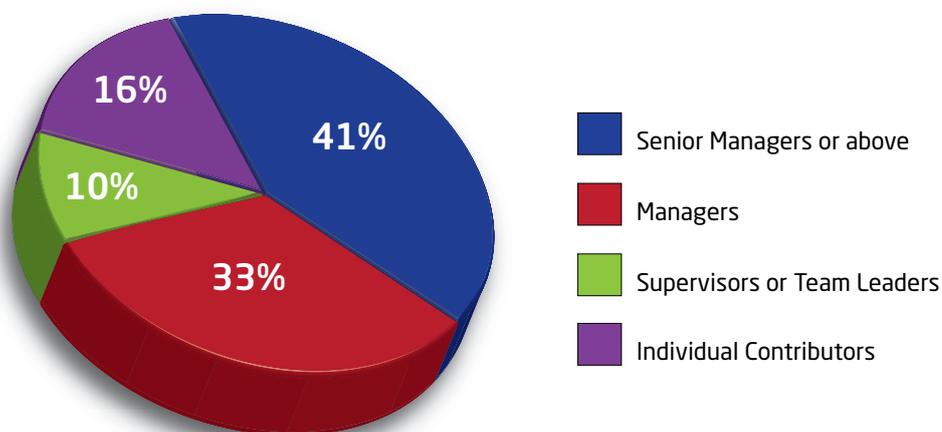
Education (highest qualification)¹²

Approximately 21% had a professional qualification, 32% had a postgraduate degree, 28% of the group had a degree, 16% were educated to high school, and the remaining 3% of the sample stated 'Not applicable' or 'other'.



Level of Responsibility¹³

50% were at senior management level (board, executive, director, vice-president and senior managers), 30% were managers, 7% were supervisors or team leaders and the remaining 13% were individual contributors.

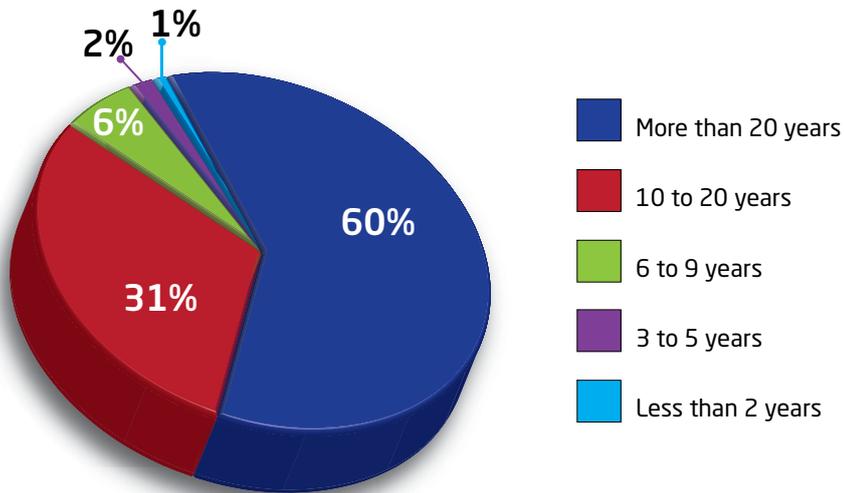


¹² Based on 64% sample response

¹³ Based on 62% sample response

Work Experience¹⁴

The overall work experience of the group ranged from 1-2 years (less than 1%) to more than 20 years (61%). 2% had between 3 and 5 years of work experience, 6% had between 6 to 9 years, and 31% had between 10 and 20 years.



¹⁴ Based on 66% sample response

Norm Group: Professionals & Managers (2011)

This sample consisted of 12592 individuals¹⁵ employed in a wide range of job areas/ functions¹⁶. Approximately 16% worked in Sales & Marketing, 10% in Human Resources, 10% in Finance, 8% in Management, 6% in Operations and 6% in Information Technology. The remaining 44% worked in other areas including Administration, Education, Law, Health, Production, Consulting, Military, Leisure, Research, Engineering, Media, Construction, Government/Civil Service and Transport.

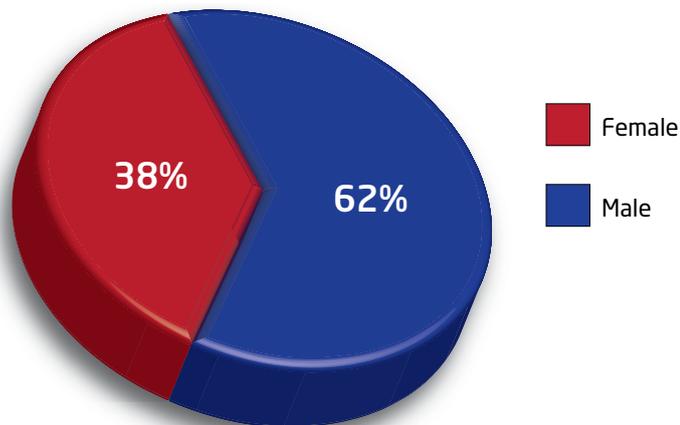
The norm group was based on a combination of all raters, including self and other raters. The breakdown for the two rater categories is provided below (with response rates for each biographical section given in the footnotes):

Self-raters

The sample consisted of 2268 professionals and managers who rated themselves on the questionnaire.

Gender¹⁷

66% of the sample were male and 34% were female.



Age¹⁸

The age of the group ranged from 24 to 67 years, with a median age of 39 years.

¹⁵ The number of participants included all raters in a given set of Performance 360 assessment (Bosses, Self, Peers and Reports). Of these, 4670 participants also completed the Performance 360 Ability Profile.

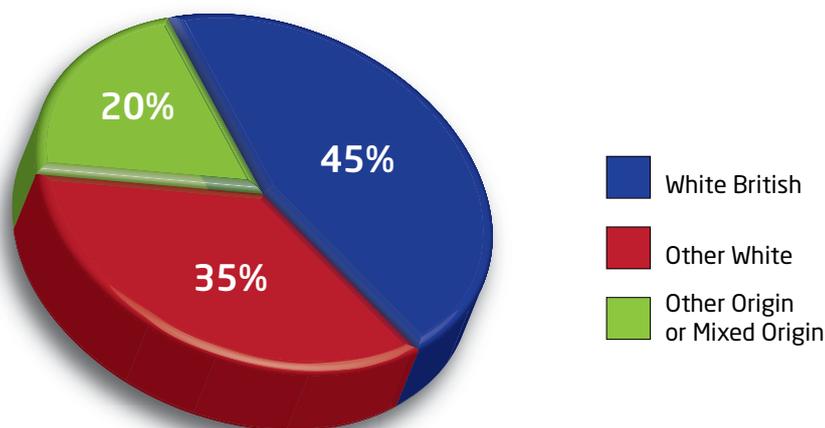
¹⁶ Based on 69% sample response

¹⁷ Based on 100% sample response

¹⁸ Based on 77% sample response

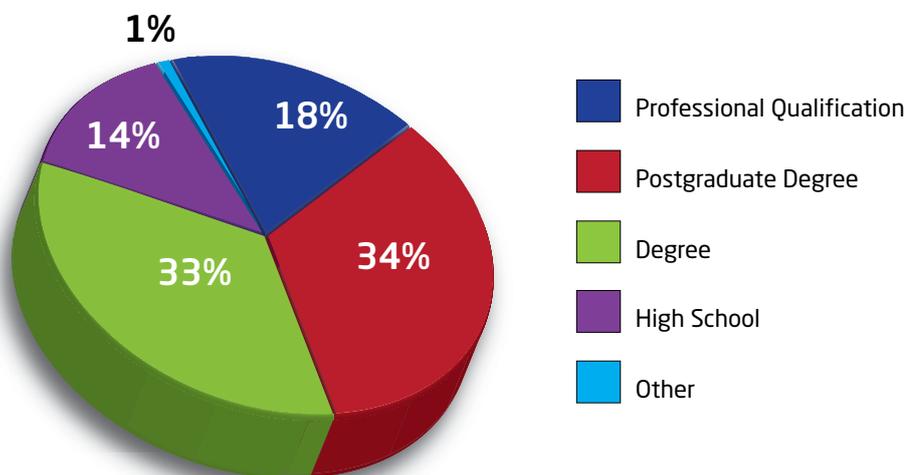
Cultural Background¹⁹

Approximately 45% of the sample described themselves as UK White and 35% as being from other White backgrounds (including European, American, Australian and New Zealander). The remaining 20% of the sample were of other (e.g., Indian, Hispanic, Black African) or mixed origin (e.g., White and Asian, White and Black African).



Education (highest qualification)²⁰

Approximately 18% had a professional qualification, 34% had a postgraduate degree, 33% of the group had a degree, 14% were educated to high school, and the remaining 1% of the group stated 'Not applicable' or 'other'.

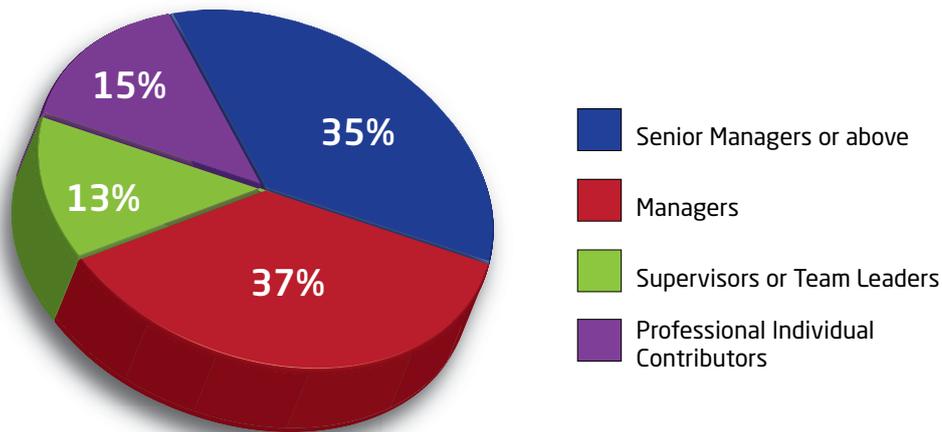


¹⁹ Based on 65% sample response

²⁰ Based on 98% sample response

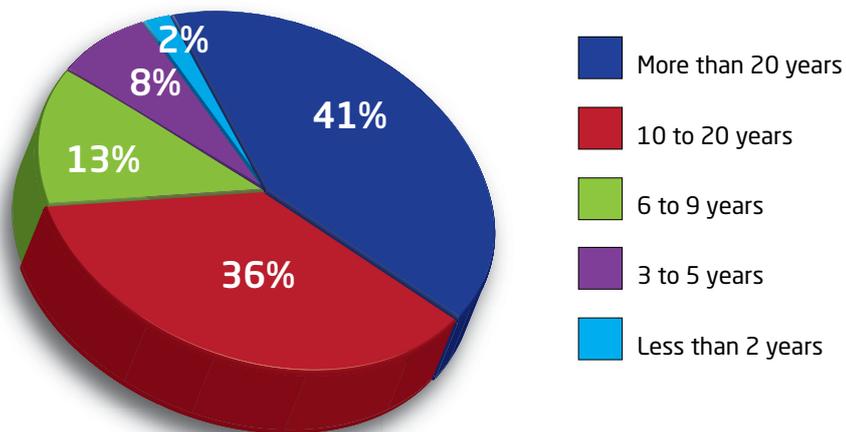
Level of Responsibility²¹

Approximately 35% of the sample were at senior management level (including board or executive level, directors, vice-presidents and senior managers), 38% were managers, 13% were supervisors or team leaders and the remaining 15% were professional individual contributors.



Work Experience²²

The overall work experience of the group ranged from 1-2 years (2%) to more than 20 years (41%). 8% had 3-5 years of work experience, 13% had between 6 and 9 years, and 36% had between 10 and 20 years.



²¹ Based on 100% sample response

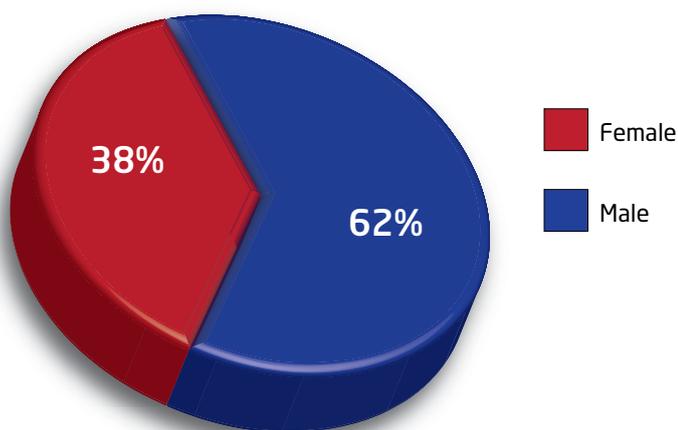
²² Based on 100% sample response

Other raters

The sample consisted of 10324 individuals who rated the work performance of a colleague in the questionnaire.

Gender²³

62% of the sample were male and 38% were female.

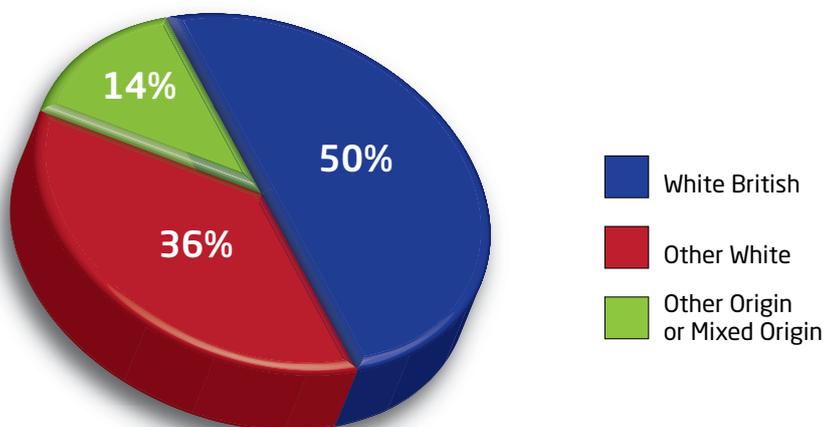


Age²⁴

The age of the group ranged from 21 to 74 years, with a median age of 42 years.

Cultural Background²⁵

Approximately 50% of the sample described themselves as UK White and 36% as being from other White backgrounds (including European, American, Australian and New Zealander). The remaining 14% of the sample were of other (e.g., Indian, Hispanic, Black African, Black American, Black Caribbean) or mixed origin (e.g., White and Asian, White and Black African, Black African and Asian).



²³ Based on 68% sample response

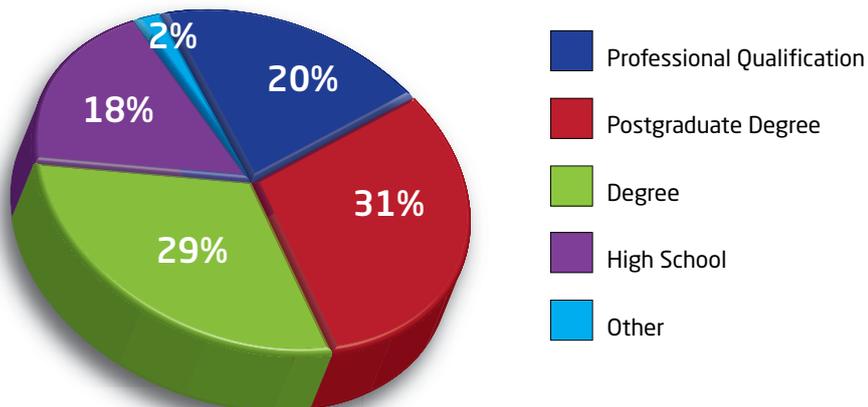
²⁴ Based on 45% sample response

²⁵ Based on 51% sample response

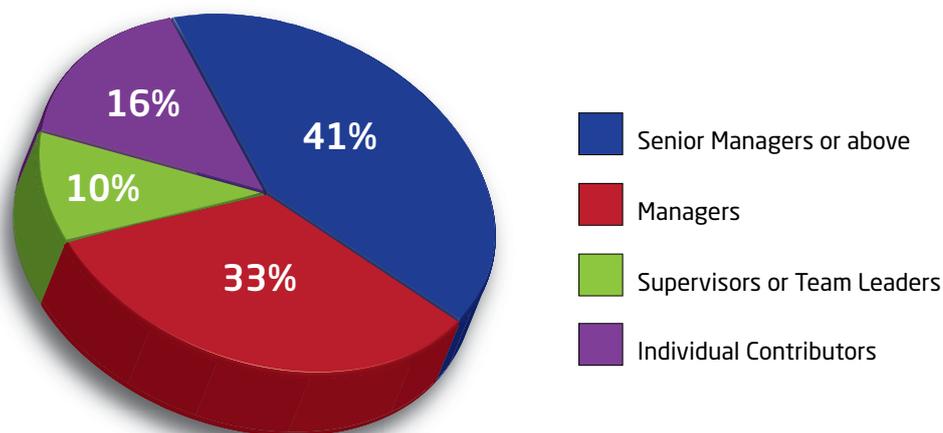
Education (highest qualification)²⁶

Approximately 20% had a professional qualification, 31% had a postgraduate degree, 29% of the group had a degree, 18% were educated to high school, and the remaining 2% of the sample stated 'Not applicable' or 'other'.

Level of Responsibility²⁷



41% were at senior management level (including board or executive level, directors, vice-presidents and senior managers), 33% were managers, 10% were supervisors or team leaders, and the remaining 16% were individual contributors (professionals and non-professionals).

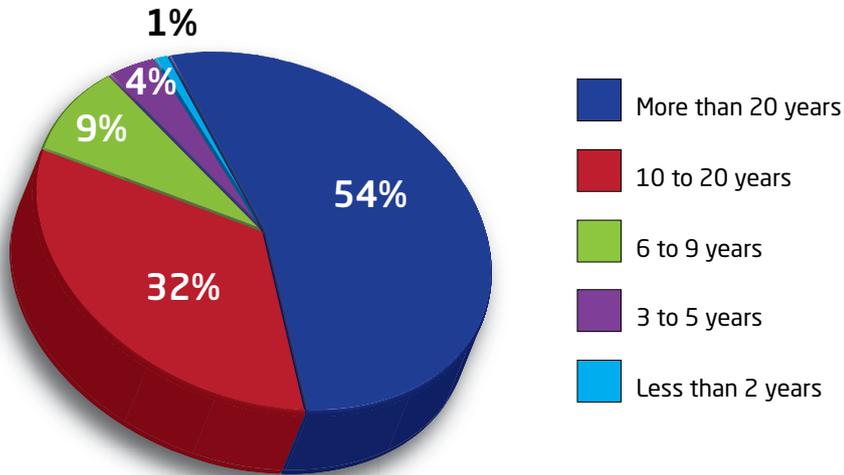


²⁶ Based on 63% sample response

²⁷ Based on 59% sample response

Work Experience²⁸

The overall work experience of the group ranged from less than 2 years (1%) to more than 20 years (54%). 4% had between 3 and 5 years of work experience, 9% had between 6 to 9 years, and 32% had between 10 and 20 years.



²⁸ Based on 65% sample response

12.3 Performance 360 (2011) Norm Tables

Table 12.2 Performance 360 Senior Managers & Executives Group Norm Table, Invited Access (N=5334)

Scale	Mean	SD
Examining Information	5.82	0.54
Documenting Facts	5.71	0.61
Interpreting Data	5.65	0.60
Developing Expertise	5.53	0.64
Adopting Practical Approaches	5.88	0.54
Providing Insights	5.69	0.64
Generating Ideas	5.43	0.76
Exploring Possibilities	5.38	0.72
Developing Strategies	5.40	0.78
Interacting with People	5.66	0.77
Establishing Rapport	5.61	0.81
Impressing People	5.02	0.87
Convincing People	5.45	0.65
Articulating Information	5.52	0.74
Challenging Ideas	5.63	0.66
Making Decisions	5.83	0.61
Directing People	5.62	0.68
Empowering Individuals	5.47	0.74
Conveying Self-Confidence	5.55	0.73
Showing Composure	5.61	0.79
Resolving Conflict	5.32	0.74
Thinking Positively	5.67	0.66
Embracing Change	5.75	0.62
Inviting Feedback	5.20	0.73
Understanding People	5.48	0.74
Team Working	5.68	0.65
Valuing Individuals	5.55	0.69
Meeting Timescales	5.60	0.79
Checking Things	5.63	0.68
Following Procedures	5.54	0.75
Managing Tasks	5.61	0.73
Upholding Standards	6.19	0.53
Producing Output	5.81	0.61
Taking Action	5.87	0.64
Seizing Opportunities	5.20	0.84
Pursuing Goals	5.74	0.65
Working with Words*	5.58	0.77
Working with Numbers*	5.69	0.85
Working with Details*	5.34	0.85
Working with Systems*	5.40	0.78
Working with Designs*	4.85	1.00
Working with Equipment*	4.72	1.14
Applying Specialist Expertise	5.67	0.63
Accomplishing Objectives	5.79	0.56
Demonstrating Potential	5.45	0.82

*This scale has a sample size of N=1564.

Table 12.3 Performance 360 Professionals & Managers Group Norm Table, Invited Access (N=12592)

Scale	Mean	SD
Examining Information	5.76	0.61
Documenting Facts	5.65	0.69
Interpreting Data	5.58	0.66
Developing Expertise	5.52	0.69
Adopting Practical Approaches	5.85	0.60
Providing Insights	5.60	0.71
Generating Ideas	5.32	0.78
Exploring Possibilities	5.30	0.74
Developing Strategies	5.22	0.83
Interacting with People	5.63	0.84
Establishing Rapport	5.67	0.86
Impressing People	4.99	0.93
Convincing People	5.35	0.73
Articulating Information	5.43	0.81
Challenging Ideas	5.51	0.74
Making Decisions	5.71	0.68
Directing People	5.49	0.78
Empowering Individuals	5.43	0.81
Conveying Self-Confidence	5.41	0.80
Showing Composure	5.53	0.85
Resolving Conflict	5.25	0.80
Thinking Positively	5.64	0.74
Embracing Change	5.68	0.69
Inviting Feedback	5.24	0.75
Understanding People	5.57	0.76
Team Working	5.70	0.68
Valuing Individuals	5.57	0.74
Meeting Timescales	5.59	0.86
Checking Things	5.61	0.75
Following Procedures	5.56	0.77
Managing Tasks	5.61	0.76
Upholding Standards	6.13	0.62
Producing Output	5.74	0.70
Taking Action	5.74	0.70
Seizing Opportunities	5.01	0.92
Pursuing Goals	5.63	0.71
Working with Words*	5.55	0.81
Working with Numbers*	5.60	0.88
Working with Details*	5.38	0.88
Working with Systems*	5.37	0.84
Working with Designs*	4.82	0.99
Working with Equipment*	4.72	1.09
Applying Specialist Expertise	5.59	0.69
Accomplishing Objectives	5.72	0.61
Demonstrating Potential	5.36	0.88

*This scale has a sample size of N=4670.

Figure 12.2 Mean Raw Scores by Senior Managers & Executives and Professionals & Managers (2011) Norm Groups

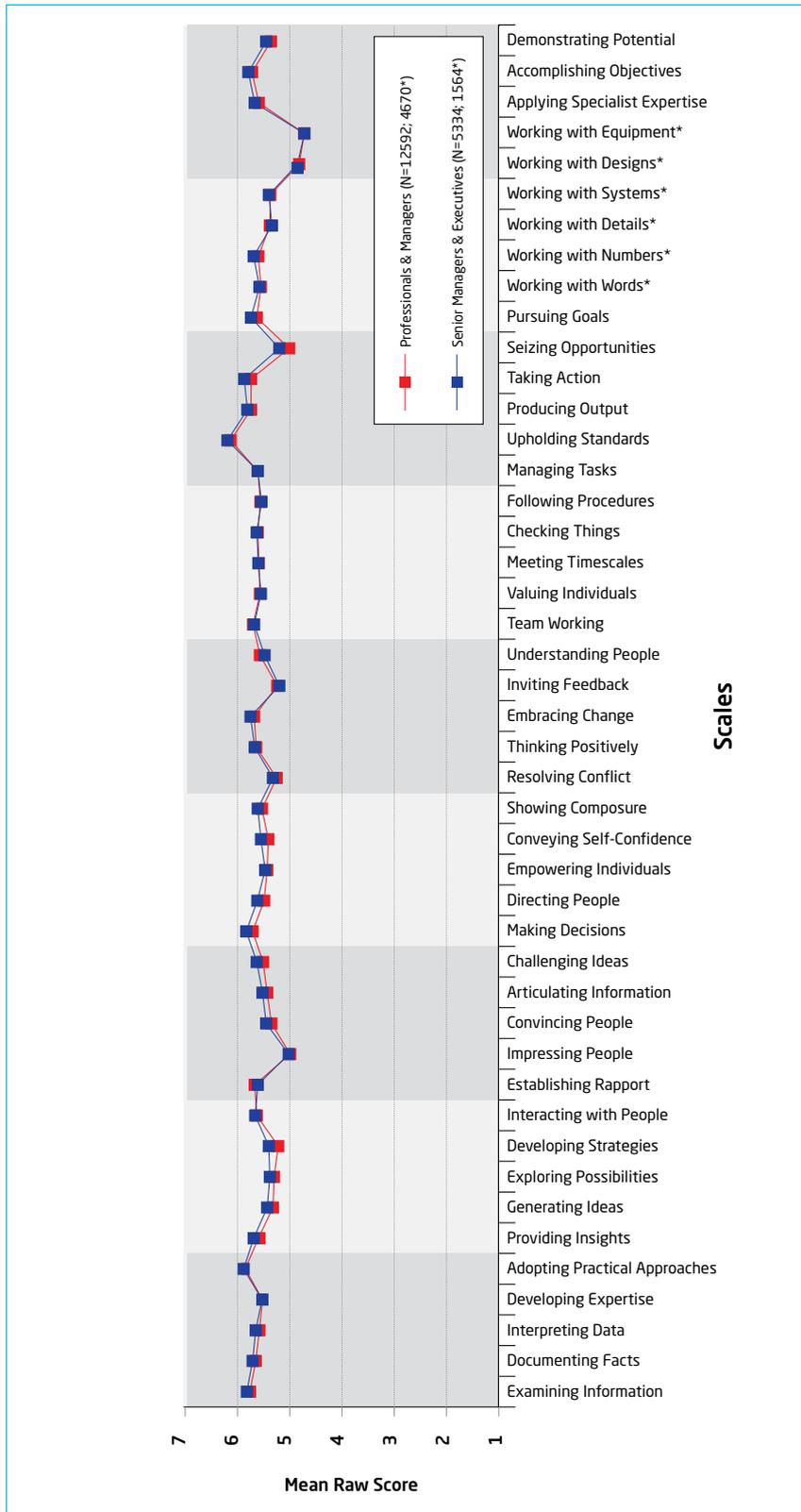


Figure 12.2 shows the mean scores for the Senior Managers & Executives (N=5334) and Professionals & Managers (N=12,592) norm groups. When the two norm groups were compared against each other, two behavioral dimensions showed a small difference (based on Cohen's d) between the groups. The Senior Managers & Executives group was rated higher than Professionals & Managers on Developing Strategies (d=.22) and Seizing Opportunities (d=.21). The two groups did not differ in other scales.

12.4 Further Reference Material

Further Information can be found in the 'Group Trends' chapter and the Appendices of this handbook.

13.0 Wave® Performance 360 Rater Trends

13.1 Background to Rater Trends

The typical rater categories in Saville Consulting Wave Performance 360 assessments are Boss, Self, Peer and Report.

The all-round perspective that can be obtained from Performance 360 allows us to study any differences in the ratings from different rater categories. Over the years, researchers have frequently found differences between categories. For example, supervisors have been found to be more lenient, and more restricted in the range of values used than peers and subordinates (Orpen, 1973). Moderate correlations have been found between subordinates and supervisors when rating on managerial performance (Mount, 1984). Similarly, subordinate ratings have been found to be in greater agreement with ratings from supervisors than self-ratings (McEvoy & Beatty, 1990). Interestingly, researchers have found that self-ratings seem to be considerably different from other raters, including supervisors and peers (Thornton, 1980; Conway, 2000).

Such disparity between ratings can, according to Murphy and DeShon (2000), be attributed to four areas of systematic difference: a) “systematic differences in what is observed”; b) “systematic differences in access to information other than observations of performance”; c) “systematic differences in expertise in interpreting what is observed”; and d) “systematic differences in evaluating what is observed”. There are such differences because bosses, peers and reports may observe or attend to different aspects of work performance (Borman, 1997), and may have different expectations of the assessee’s performance (cf. Tubré et al., 2006; Conway, 2000; Fletcher & Baldry, 1999; Greguras & Robie, 1998; Garvin, 1995). Some raters are also likely to observe more of the assessee’s work behaviors than others (Heneman, Wexley, & Moore, 1987). Murphy and DeShon propose that these are differences rather than random measurement error.

The variance which we see between ratings from the different categories may reflect the different types of relationships between raters and the assessee. Scullen, Mount and Judge (2003) have suggested, for example, that while bosses are able to rely on formal authority to secure required resources from a subordinate, peers are more likely to depend on interpersonal relationships. As a result, peer ratings may place greater emphasis than boss ratings on the assessee’s ability to work in a team and cooperate with others. The studies of Conway (1999) and Fox & Bizman (1988) support the notion that raters are more likely to consider interpersonal relationships when rating their peers than when rating subordinates. Subordinates’ ratings of their bosses may, according to Scullen, Mount and Judge (2003), be influenced by yet another set of preoccupations. Leadership skills and fairness may be of particular importance to subordinates, as may their bosses’ ability and willingness to help them develop their technical and administrative competence. This assertion is supported by Fox and Bizman’s (1988) study, which found that raters, when asked to evaluate their bosses’ performance, were particularly likely to focus on managerial abilities, interpersonal relations, and professional knowledge.

13.2 Performance 360 Rater Trends

Figures 13.1 and 13.2 show the mean ratings, in raw score and sten respectively, of our Wave Performance 360 data, in terms of the four typical rater categories Boss, Self, Peer and Report. The data contain 1320 sets of Performance 360 assessment. Each assessment set contains ratings from self and at least one boss, two peers and two reports. An aggregated score was calculated for each rater category within an assessment. The mean ratings are the average of aggregate scores across all assessments. As the Wave Performance 360 questionnaire is based on the Wave model, raters from all categories are required to evaluate the same aspects of the assessee's work performance, covering Solving Problems, Influencing People, Adapting Approaches, Delivering Results and Overall Performance (some may also cover the additional area of Ability). It can be seen from the graphs that the ratings from the different rater categories are similar. Therefore, for simplicity, results presented on a Performance 360 Report are benchmarked against only one norm group which is derived from the average of the four rater categories, instead of separate norm groups for each rater category.

Table 13.1 shows the inter-correlations of ratings between the four rater categories, based on the aggregate scores in the same data set. Overall, the highest average correlation is between Boss and Peer ($r=.36$), followed by the correlation between Peer and Report ($r=.35$) then between Boss and Report ($r=.27$). The correlations between self-ratings and other raters' are generally lower, ranging from .19 (Boss) to .21 (Peer). These results support previous findings obtained by Thornton (1980) and Conway (2000) that self-ratings seem to be more distinct from other raters'. Two behavior dimensions, Adopting Practical Approaches and Inviting Feedback, show particularly low agreement between Self and the other three rater categories.

Results from Table 13.1 also show that the agreements between rater categories are higher in some behavior constructs than others. At cluster level, higher agreement is found in Influencing People and Delivering Results. Influencing People looks at how one influences and works with other people, and Delivering Results concerns one's approach to implementing and achieving results. It may be that these constructs are more observable. The specific constructs under the two clusters that show higher agreement between categories are Interacting with People, Establishing Rapport, Articulating Information and Meeting Timescales.

In general, the biggest disconnect between raters' views is between self and all other raters. And after that, the two rater groups most likely to disagree on the assessee are, perhaps unsurprisingly, bosses and reports.

Figure 13.1 Mean Raw Scores by Rater Category

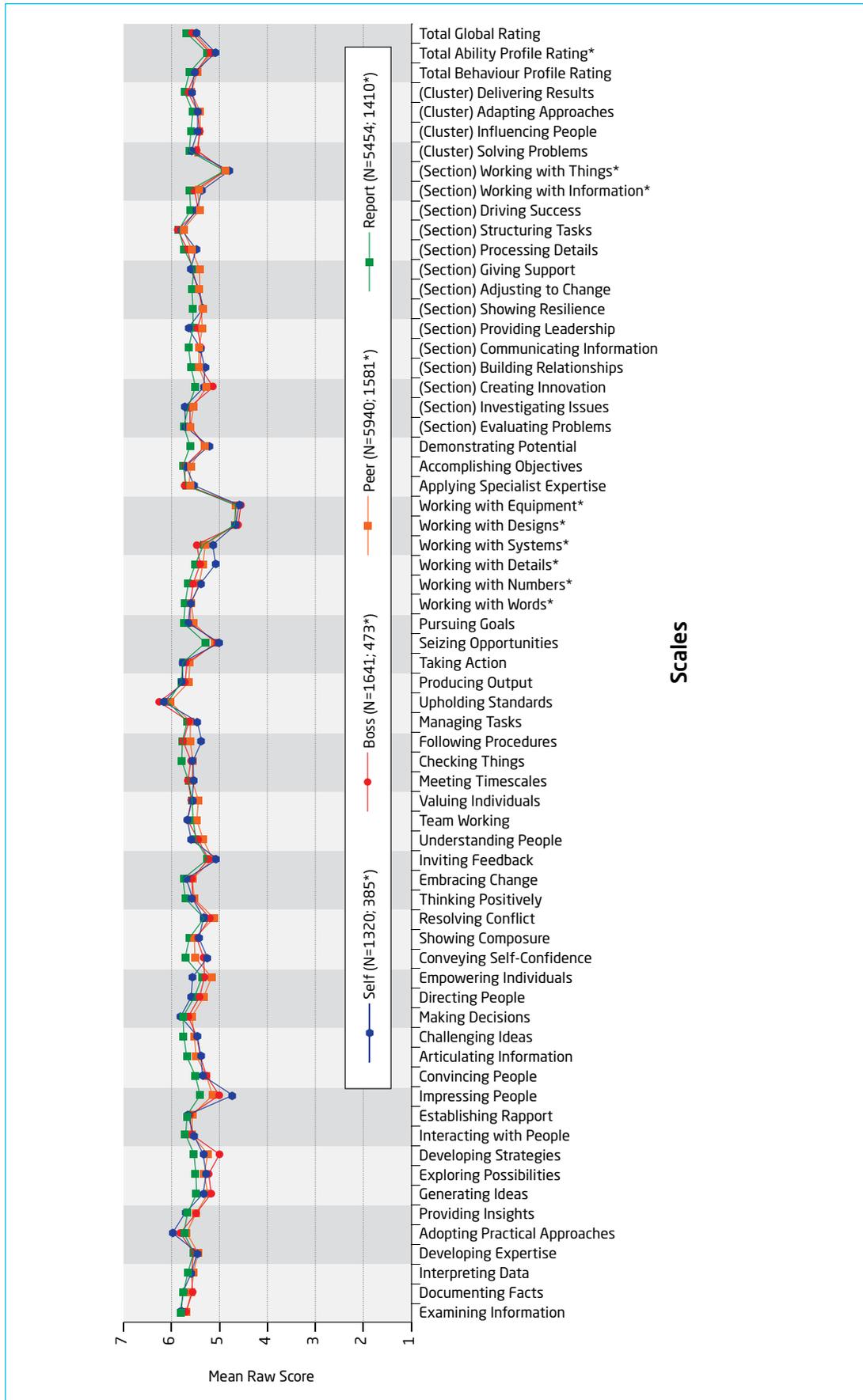


Figure 13.2 Mean Sten Scores by Rater Category

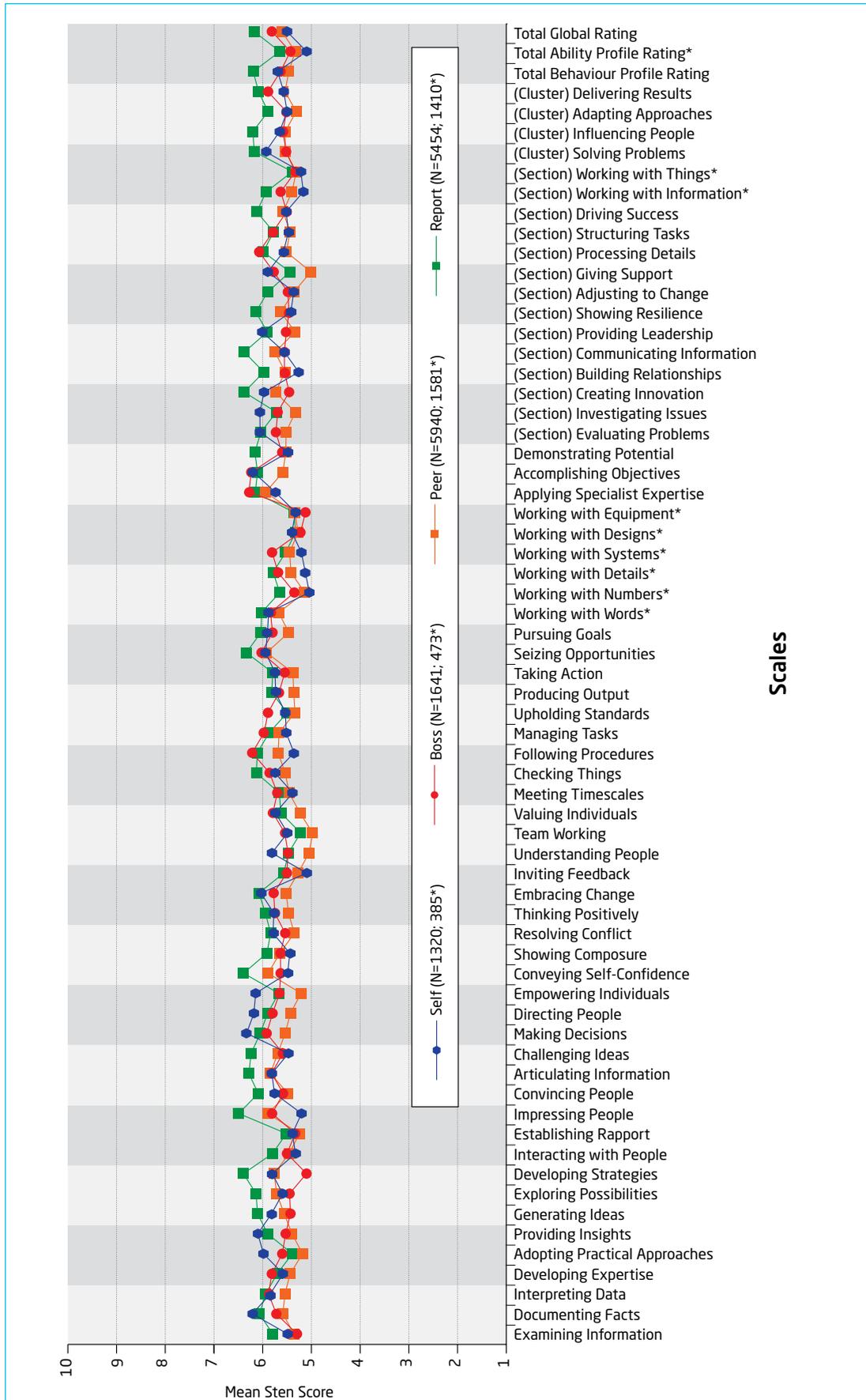


Table 13.1 Inter-correlations of Wave Performance 360 Scales between Rater Categories (Sets of assessment N=1320) - Continued on next page

Scale	Self v Boss	Self v Peer	Self v Report	Boss v Peer	Boss v Report	Peer v Report
Examining Information	.14	.10	.12	.31	.23	.32
Documenting Facts	.19	.22	.19	.35	.25	.30
Interpreting Data	.17	.15	.13	.31	.21	.28
Developing Expertise	.13	.21	.20	.26	.25	.31
Adopting Practical Approaches	.05	.09	.12	.24	.18	.27
Providing Insights	.13	.15	.09	.27	.22	.29
Generating Ideas	.22	.20	.14	.37	.26	.33
Exploring Possibilities	.16	.17	.10	.33	.23	.25
Developing Strategies	.25	.23	.18	.34	.25	.30
Interacting with People	.32	.32	.26	.43	.37	.42
Establishing Rapport	.32	.34	.29	.50	.41	.45
Impressing People	.16	.21	.21	.38	.28	.32
Convincing People	.15	.15	.15	.31	.29	.34
Articulating Information	.22	.27	.21	.36	.29	.41
Challenging Ideas	.15	.18	.12	.26	.19	.29
Making Decisions	.17	.17	.18	.29	.24	.34
Directing People	.17	.17	.18	.35	.32	.38
Empowering Individuals	.18	.23	.23	.37	.32	.40
Conveying Self-Confidence	.25	.20	.18	.39	.29	.33
Showing Composure	.19	.25	.27	.40	.33	.41
Resolving Conflict	.16	.19	.25	.36	.28	.32
Thinking Positively	.18	.22	.21	.35	.28	.31
Embracing Change	.18	.16	.13	.28	.19	.29
Inviting Feedback	.02	.03	.05	.22	.15	.25
Understanding People	.18	.22	.22	.37	.30	.39
Team Working	.12	.12	.17	.32	.25	.34
Valuing Individuals	.14	.18	.20	.30	.26	.32
Meeting Timescales	.34	.31	.30	.48	.34	.44
Checking Things	.21	.20	.23	.38	.28	.34
Following Procedures	.21	.27	.27	.39	.24	.36
Managing Tasks	.24	.22	.23	.39	.29	.38
Upholding Standards	.15	.14	.15	.35	.30	.35
Producing Output	.15	.20	.20	.36	.29	.38
Taking Action	.22	.22	.20	.37	.33	.38
Seizing Opportunities	.28	.35	.30	.37	.31	.38
Pursuing Goals	.19	.20	.17	.34	.27	.35
Working with Words*	.24	.34	.32	.36	.28	.34
Working with Numbers*	.27	.35	.34	.46	.33	.41
Working with Details*	.15	.15	.25	.32	.21	.23
Working with Systems*	.25	.27	.25	.31	.20	.30
Working with Designs*	.30	.31	.29	.27	.17	.30
Working with Equipment*	.33	.36	.31	.34	.23	.38
Applying Specialist Expertise	.17	.20	.11	.34	.22	.32
Accomplishing Objectives	.11	.13	.13	.32	.21	.29
Demonstrating Potential	.22	.27	.18	.41	.31	.38

*This scale is based on 385 sets of assessment.

Table 13.1 Inter-correlations of Wave Performance 360 Scales between Rater Categories (Sets of assessment N=1320) - Continued

Score	Self v Boss	Self v Peer	Self v Report	Boss v Peer	Boss v Report	Peer v Report
(Section) Evaluating Problems	.18	.15	.14	.36	.26	.32
(Section) Investigating Issues	.08	.13	.12	.31	.25	.30
(Section) Creating Innovation	.24	.22	.16	.40	.28	.33
(Section) Building Relationships	.30	.31	.26	.50	.41	.44
(Section) Communicating Information	.17	.19	.14	.35	.28	.37
(Section) Providing Leadership	.18	.17	.19	.38	.33	.41
(Section) Showing Resilience	.18	.17	.21	.40	.32	.36
(Section) Adjusting to Change	.11	.12	.11	.31	.22	.30
(Section) Giving Support	.17	.20	.22	.38	.31	.39
(Section) Processing Details	.28	.29	.28	.48	.32	.41
(Section) Structuring Tasks	.19	.17	.19	.40	.32	.40
(Section) Driving Success	.26	.29	.24	.40	.34	.39
(Section) Working with Information*	.18	.23	.27	.39	.25	.31
(Section) Working with Things*	.34	.35	.34	.33	.24	.36
(Cluster) Solving Problems	.16	.15	.13	.39	.28	.34
(Cluster) Influencing People	.20	.21	.19	.44	.38	.42
(Cluster) Adapting Approaches	.13	.14	.18	.39	.31	.37
(Cluster) Delivering Results	.21	.18	.20	.44	.33	.39
Total Behavior Profile Rating	.11	.09	.13	.40	.33	.38
Total Ability Profile Rating*	.25	.25	.30	.33	.21	.30
Total Global Rating	.15	.17	.12	.39	.29	.36
Mean	.19	.21	.20	.36	.27	.35
Median	.18	.20	.19	.36	.28	.34
Min	.02	.03	.05	.22	.15	.23
Max	.34	.36	.34	.50	.41	.45

*This scale is based on 385 sets of assessment.

14.0 Reliability

When people are assessed on different occasions or on different versions of the same test, why do they get different scores? It is because we cannot measure people's characteristics or traits with perfect reliability.

Reliability estimates provide an indication of how precise and error-free an assessment is in measuring the desired characteristics or traits.

The reliability of an assessment is an important prerequisite in allowing the assessment user to draw correct inferences from the assessment scores. The observed scores on the assessment are intended to provide an approximation of the individual's true scores. If test or profile scores are unreliable, they provide a less precise and less accurate reflection of the individual's true scores. The higher the reliability, the less the error and the more likely the observed scores are an accurate reflection of the individual's true scores.

Reliability is merely a stepping stone towards, or a prerequisite of, test or questionnaire validity, which we discuss in more detail in the next chapter. If a test user is to draw correct and meaningful inferences from assessment scores, then the assessment must be reliable. But in order for an assessment to be used accurately and fairly in a workplace context, it must also be supported by appropriate validity data. The greater the reliability, the greater the chance of high validity.

There are several methods of estimating test reliability; four common approaches are detailed below.

Internal Consistency Reliability

This form of reliability is an index of how the items in an assessment (e.g., a personality scale) relate to one another. It carries the practical advantage that it can be estimated without the need for a retest or an alternate form, but there are some drawbacks.

For questionnaires, it is important that internal consistency reliability is satisfactorily high without being artificially inflated. For instance, a personality scale with repetitive item content will have high internal consistency reliability estimates but may lack breadth of measurement. This narrow coverage of the content domain in a questionnaire may fall short of what scales should be measuring, and is likely to impact on the empirical validity of the test in forecasting effectiveness on independently assessed criteria.

A development aim of Wave® Performance 360 was to have moderate internal consistency at Section level, with reliability coefficients between .60 and .90. In essence, this form of reliability was seen by the authors as a measure of the breadth or narrowness of the scale.

Test-Retest Reliability

One key means of estimating reliability is to look at the stability of test scores over time. This can be accomplished by a group of individuals completing the same assessment on two separate occasions.

The Pearson Product-Moment correlation coefficient provides this estimate of reliability by comparing scores obtained on the first occasion with those obtained on the second occasion.

Alternate Form Reliability

Where two or more versions of an assessment have been developed, it is possible to estimate the reliability between the versions. This, for example, has been established for Wave Professional Styles and Focus Styles, but is not possible for Wave Performance 360 as there is only one version. However, while there is, strictly speaking, no alternate form for Performance 360, there is an in-depth version of the assessment. Information on the study of the relationship between Performance 360 and this in-depth measure can be found in the 'Validity' chapter of this handbook.

Inter-Rater Reliability

This form of reliability is estimated by having two individuals rate the performance of one person they know (assessee), using the same questionnaire. The correlation of the ratings on a particular dimension between the two raters provides one estimate of the inter-rater reliability.

Internal consistency reliability, test-retest reliability and alternate form reliability are forms of intra-rater reliability indices, while inter-rater reliability is the only form that looks at the reliability of ratings between raters.

Some Potential Sources of Error

Assessment scores can contain errors of measurement from a number of sources, for example:

- **Questionnaire Design** - questions with negative phrasing or those touching on more than one aspect in an item tend to increase measurement error
- **Individual** - mood, temperament, motivation, well-being
- **Environment** - noise, temperature, presence of others
- **Administration** - degree and consistency of standardization
- **Scoring** - the accuracy of the scoring key and scoring process

Some Potential Sources of Error in Different Ratings

In multi-rater assessments there are many different sources of variance in ratings and the interactions between different sources can be complicated. According to Murphy & DeShon (2000), sometimes variance in ratings may be idiosyncratic, but some systematic factors including rater characteristics (e.g., leniency) and the context in which the rater operates (e.g., relationship with the assessee - boss, peer or report, rater's position in the company) can lead to rating variance. The researchers commented further that even when there are similarities between different raters (e.g., shared goals and biases, shared perceptions of the organization and of appraisal systems, shared frames of reference and shared relationships with the assessee), there can still be agreement between their ratings but without 'true score' variance in them. These variances are not results of the assessee's true performance but can correlate with it.

Over time, errors can be looked at in a more systematic way to uncover the different sources of variance. One approach is to use generalizability studies. Based on analysis of variance, such studies take into account both systematic and unsystematic sources of variance and differentiate them when evaluating the reliability of an assessment (Cronbach et al., 1963).

In this chapter, we look at the inter-rater reliability of Wave Performance 360 using intraclass correlations. Correlational analyses are based on Classical Test Theory, which states that observed scores equal true scores plus error. In Classical Test Theory, multiple measurement errors are treated as a unitary construct. As the main focus of this chapter is not on identifying all the different sources of rating variance but the reliability of the Performance 360 assessment, errors are seen as a whole and, therefore, correlations are used. Researchers also agree that intraclass correlation coefficients are analogous to the coefficients of generalizability (c.f. Cronbach et al., 1964).

Standard Error of Measurement (SEm)

When test or assessment users receive a test score, they make inferences, communicate and/or make decisions based on the test score. However, the observed score is subject to error, and in order to be in a better position to use the test score, it is important for a test user to have an appreciation of the band of error around the score, and to know how likely it is to contain the individual's true score. To do this, the Standard Error of Measurement is computed.

Formula

The Standard Error of Measurement equals the Standard Deviation of a group multiplied by the square root of one minus the reliability coefficient.

$$SEm = SD \sqrt{1 - r_t}$$

Where:

- SEm = Standard Error of Measurement
- SD = Standard Deviation of the sample that the reliability coefficient was calculated from
- r_t = the reliability coefficient (test-retest, alternate form, internal consistency)

If we take the average alternate reliability of the Wave Styles scales, which was $r = .86$ in the standardization trials, and want to calculate the Standard Error of Measurement for a sten score, then:

SD Sten Score = 2

Alternate Form Reliability = .86

$$\begin{aligned} \text{SEm} &= 2 \sqrt{1 - 0.86} \\ &= 2 \times .37 \\ &= .74 \end{aligned}$$

A band of 1 SEm (.74 stens) either side of an individual's score results in a 68% probability that this band contains the true score for the individual. For instance, with a sten score of 6, we are confident that 68% of the time the person's true score will be between 5.26 and 6.78 - or within 1 SEm to either side of the observed score.

Placing a band of 2 SEms (i.e., $2 \times .74$ stens or 1.48 stens) either side of the observed score gives a 96% probability that this band contains the true score for this individual

+/- 1SEm - 68% Probability

+/- 2SEm - 96% Probability

14.1 Reliability Overview

Internal Consistency Reliability

Table 14.1 provides the Internal Consistency Reliability (Cronbach's Alpha) of the Wave Performance 360 scales at a higher level. These higher level scales of Wave Performance 360 were designed to have moderate internal consistency, with an objective of obtaining estimates ranging from a minimum of .60 to a maximum of .90. The table shows that the internal consistency reliability coefficients of all the scales are within this desired range, with the lowest at .69 (Showing Resilience and Working with Information) and highest at .84 (Giving Support).

Table 14.1 Internal Consistency Reliability of Wave Performance 360 higher level scales (N=25,135)

Scale	Mean	SD	SEm (Sten)	r
Behavior Sections				
Evaluating Problems	16.96	2.65	.77	.85
Investigating Issues	16.80	2.57	.96	.77
Creating Innovation	15.88	2.91	.80	.84
Building Relationships	16.43	3.18	.89	.80
Communicating Information	16.37	2.74	.96	.77
Providing Leadership	16.28	3.01	.87	.81
Showing Resilience	16.24	2.77	1.11	.69
Adjusting Change	16.45	2.71	1.00	.75
Giving Support	16.58	3.10	.72	.87
Processing Details	16.86	2.83	.89	.80
Structuring Tasks	17.41	2.55	1.00	.75
Driving Success	16.37	2.78	.87	.81
Ability Sections				
Working with Information	16.53	2.62	1.06	.72
Working with Things	14.75	2.85	.89	.80
Global Cluster				
Global (Overall) Effectiveness	16.66	2.62	1.00	.75
Mean	16.44	2.79	.92	.79
Median	16.45	2.77	.89	.80
Min	14.75	2.55	.72	.69
Max	17.41	3.18	1.11	.87

Test-Retest Reliability

Tables 14.2 and 14.3 provide the test-retest reliability of Wave Performance 360 scales at both higher and lower levels. The Performance 360 questionnaire was completed online by an external rater (N=142) assessing an individual. It was readministered online to the same external raters who were rating the same individual with a 6-month interval between the two administrations.

The test-retest reliabilities for scales at a higher level (Table 14.2) range from .29 (Providing Leadership) to .53 (Processing Details), with an average of .41. Table 14.3 shows that the average test-retest reliability for scales at a lower level is .36. The behavior dimension that shows lowest test-retest reliability is Directing People (.25), one of the three dimensions under the behavior section Providing Leadership. This is followed by Developing Expertise (.26), Interacting with People (.26) and Accomplishing Objectives (.26). The behavior dimension with highest test-retest reliability is Meeting Timescales (.58), which falls under the Processing Details behavior section. The other dimensions which show higher reliabilities are Working with Words (.49), Seizing Opportunities (.47), Taking Action (.46) and Pursuing Goals (.46).

Overall, these results indicate that ratings of performance by others on Performance 360 are less stable over time than, for example, self-report of styles (see *Wave Professional Styles Handbook*). It should be noted that these are test-retests of individual raters rather than of overall aggregation across a number of raters. The values of the test-retest for these aggregated scores are expected to be higher. Nevertheless, the data indicate that we should not expect Performance 360 ratings of effectiveness to be highly stable over time.

Table 14.2 External Rater Test-Retest Reliability of Wave Performance 360 higher level scales (N=142)

Scale	Mean _{t1}	SD _{t1}	Mean _{t2}	SD _{t2}	SEm (Sten)	r
Behavior Sections						
Evaluating Problems	16.45	2.82	16.39	3.05	1.55	.40
Investigating Issues	16.68	2.60	16.62	2.89	1.59	.36
Creating Innovation	15.05	2.97	15.12	3.29	1.47	.46
Building Relationships	16.63	3.17	16.75	2.99	1.66	.31
Communicating Information	15.73	2.96	15.61	3.29	1.49	.44
Providing Leadership	15.82	3.14	15.96	3.02	1.68	.29
Showing Resilience	15.73	2.77	15.64	3.18	1.51	.43
Adjusting to Change	15.51	3.09	15.55	3.04	1.50	.44
Giving Support	16.36	3.30	16.56	3.06	1.58	.38
Processing Details	16.80	2.74	16.92	2.88	1.37	.53
Structuring Tasks	17.39	2.28	17.38	2.64	1.45	.47
Driving Success	15.82	2.96	15.93	3.07	1.43	.49
Ability Sections						
Working with Information	15.65	2.91	15.51	2.93	1.44	.48
Working with Things	14.10	3.02	13.43	3.28	1.64	.33
Global Cluster						
Global (Overall) Effectiveness	16.61	2.73	16.43	2.87	1.61	.35
Mean	16.02	2.90	15.99	3.03	1.53	.41
Median	15.82	2.96	15.96	3.04	1.51	.43
Min	14.10	2.28	13.43	2.64	1.37	.29
Max	17.39	3.30	17.38	3.29	1.68	.53

Note: Subjects rated the same assessee twice at a 6 month interval.

Table 14.3 External Rater Test-Retest Reliability of Wave Performance 360 lower level scales (N=142)

Scale	Mean _{t1}	SD _{t1}	Mean _{t2}	SD _{t2}	SEm (Sten)	r
Behavior Dimensions						
Examining Information	5.60	.99	5.63	1.11	1.69	.29
Documenting Facts	5.46	1.22	5.56	1.24	1.67	.30
Interpreting Data	5.39	1.07	5.20	1.14	1.51	.43
Developing Expertise	5.56	1.11	5.56	1.11	1.72	.26
Adopting Practical Approaches	5.68	.99	5.71	1.10	1.69	.29
Providing Insights	5.44	1.08	5.35	1.18	1.61	.35
Generating Ideas	5.08	1.15	5.12	1.21	1.64	.33
Exploring Possibilities	5.01	1.20	5.11	1.25	1.51	.43
Developing Strategies	4.95	1.15	4.89	1.34	1.61	.35
Interacting with People	5.60	1.20	5.70	1.16	1.72	.26
Establishing Rapport	5.82	1.19	5.86	1.10	1.62	.34
Impressing People	5.20	1.34	5.19	1.37	1.60	.36
Convincing People	5.20	1.06	5.22	1.16	1.55	.40
Articulating Information	5.16	1.31	5.15	1.33	1.61	.35
Challenging Ideas	5.36	1.14	5.25	1.33	1.59	.37
Making Decisions	5.44	1.11	5.45	1.12	1.66	.31
Directing People	5.18	1.26	5.20	1.27	1.73	.25
Empowering Individuals	5.20	1.31	5.31	1.24	1.69	.29
Conveying Self Confidence	5.37	1.09	5.38	1.28	1.60	.36
Showing Composure	5.16	1.39	5.08	1.41	1.54	.41
Resolving Conflict	5.20	1.22	5.18	1.32	1.62	.34
Thinking Positively	5.32	1.25	5.33	1.21	1.54	.41
Embracing Change	5.30	1.15	5.30	1.20	1.59	.37
Inviting Feedback	4.89	1.28	4.92	1.28	1.54	.41
Understanding People	5.44	1.35	5.49	1.25	1.54	.41
Team Working	5.42	1.22	5.55	1.16	1.71	.27
Valuing Individuals	5.51	1.13	5.51	1.17	1.64	.33
Meeting Timescales	5.57	1.33	5.66	1.20	1.30	.58
Checking Things	5.61	1.03	5.64	1.09	1.61	.35
Following Procedures	5.63	1.07	5.61	1.12	1.48	.45
Managing Tasks	5.61	.97	5.55	1.15	1.52	.42
Upholding Standards	5.96	.91	6.04	1.00	1.59	.37
Producing Output	5.82	.98	5.79	1.10	1.56	.39
Taking Action	5.46	1.13	5.49	1.01	1.47	.46
Seizing Opportunities	4.85	1.36	4.96	1.36	1.46	.47
Pursuing Goals	5.50	1.08	5.47	1.17	1.47	.46
Ability Dimensions						
Working with Words	5.49	1.15	5.50	1.11	1.43	.49
Working with Numbers	5.07	1.29	4.95	1.26	1.54	.41
Working with Details	5.28	1.31	5.10	1.30	1.55	.40
Working with Systems	5.02	1.26	4.93	1.25	1.65	.32
Working with Designs	4.60	1.22	4.41	1.25	1.71	.27
Working with Equipment	4.56	1.38	4.33	1.38	1.69	.29
Global Sections						
Applying Specialist Expertise	5.64	1.01	5.56	1.10	1.67	.30
Accomplishing Objectives	5.78	.95	5.50	1.03	1.72	.26
Demonstrating Potential	5.35	1.32	5.39	1.38	1.61	.35
Mean	5.35	1.17	5.34	1.21	1.59	.36
Median	5.39	1.15	5.38	1.20	1.61	.35
Min	4.56	.91	4.33	1.00	1.30	.25
Max	5.96	1.39	6.04	1.41	1.73	.58

Note: Subjects rated the same assessee twice at a 6 month interval.

Inter-Rater Reliability

Intraclass correlations¹ were conducted to look at the reliability of raters within rater categories with multiple raters (i.e., peers and reports).

Tables 14.4 and 14.5 show the intraclass correlations for peer and report categories for Wave Performance 360 scales at higher and lower levels respectively. The analysis included Performance 360 assessments which had 3 peers and those with 3 reports. The tables show the reliability of the average or sum of the ratings by the 3 peers or reports. The reliability of the peers' ratings were .49 at the higher level and .45 at the lower level, and the reliability of the reports' ratings were .48 at the higher level and .44 at the lower level. The results provide evidence of the inter-rater reliability of Performance 360 where there are 3 raters in a category. Clearly, although there are differences in perceptions between raters, there is a degree of agreement between them overall.

Table 14.4 Inter-Rater Reliability - Intraclass Correlations of the Peer and Report categories for Wave Performance 360 higher level scales

Scale	Peer (N=4,554)	Report (N=3,603)
Behavior Sections		
Evaluating Problems	.43	.42
Investigating Issues	.47	.43
Creating Innovation	.47	.46
Building Relationships	.58	.56
Communicating Information	.49	.48
Providing Leadership	.52	.56
Showing Resilience	.44	.51
Adjusting to Change	.43	.44
Giving Support	.51	.56
Processing Details	.56	.53
Structuring Tasks	.50	.49
Driving Success	.53	.53
Ability Sections		
Working with Information	.45	.48
Working with Things	.52	.31
Global Cluster		
Global (Overall) Effectiveness	.49	.45
Mean	.49	.48
SD	.04	.07
Max	.58	.56
Min	.43	.31

¹The one-way random effects model was used in these intraclass correlations. Such a model is to be used when it is not possible to associate any rating variability with a particular rater. In 360 assessment datasets, assessees tend to have different raters and different numbers of raters. It is, therefore, difficult to clearly separate the sources of variability, which can come from specific raters, interactions of raters with the assessees as well as measurement error of the assessment.

Table 14.5 Inter-Rater Reliability - Intraclass Correlations of the Peer and Report categories for Wave Performance 360 lower level scales

Scale	Peer (N=4,554)	Report (N=3,603)
Behavior Dimensions		
Examining Information	.42	.42
Documenting Facts	.36	.40
Interpreting Data	.37	.38
Developing Expertise	.44	.34
Adopting Practical Approaches	.39	.38
Providing Insights	.41	.43
Generating Ideas	.44	.41
Exploring Possibilities	.38	.38
Developing Strategies	.43	.41
Interacting with People	.54	.52
Establishing Rapport	.56	.58
Impressing People	.45	.46
Convincing People	.42	.44
Articulating Information	.47	.48
Challenging Ideas	.40	.42
Making Decisions	.40	.51
Directing People	.48	.53
Empowering Individuals	.48	.50
Conveying Self-Confidence	.39	.38
Showing Composure	.48	.55
Resolving Conflict	.42	.46
Thinking Positively	.48	.42
Embracing Change	.36	.37
Inviting Feedback	.33	.41
Understanding People	.51	.52
Team Working	.40	.49
Valuing Individuals	.43	.49
Meeting Timescales	.53	.57
Checking Things	.46	.47
Following Procedures	.48	.45
Managing Tasks	.46	.44
Upholding Standards	.47	.43
Producing Output	.43	.49
Taking Action	.43	.44
Seizing Opportunities	.54	.50
Pursuing Goals	.44	.49
Ability Dimensions		
Working with Words	.43	.49
Working with Numbers	.52	.55
Working with Details	.42	.43
Working with Systems	.45	.25
Working with Designs	.49	.24
Working with Equipment	.46	.33
Global Sections		
Applying Specialist Expertise	.44	.40
Accomplishing Objectives	.41	.42
Demonstrating Potential	.52	.45
Mean	.45	.44
SD	.05	.07
Max	.56	.58
Min	.33	.24

14.2 Summary of Reliability

In this chapter, we have assessed the reliability of Wave Performance 360 through intra-rater reliability (internal consistency reliability and test-retest reliability) and inter-rater reliability (intraclass correlation) measures. Internal consistency estimates of Wave Performance 360 indicate that there is convergence between the items within each of the higher level scales in the assessment. The average internal consistency was .79. Clearly, Wave Performance 360 is not a measure of personality or styles but provides information on the evaluation given by different raters on the work performance of one person. Such performance evaluations by others are less stable over time than self-reports of styles or personality (the average test-retest reliability for external raters was .41 for higher level scales) and there is only a moderate degree of agreement between raters as to the effectiveness of individuals they are rating (average intraclass correlations for Peers and Reports were .49 and .48 respectively for higher level scales).

It is important to recognize that in 360 assessments a number of features may impact ratings from different perspectives.

Firstly, the type of rating scale the questionnaire adopts can be important. Wave Performance 360 uses a rating scale that has purposely been chosen to address effectiveness. Measuring this area calls for an evaluative judgment by the rater on whether the behaviors they see are effective or ineffective. The authors regard this as an important feature in assessing performance as opposed to other questionnaires which may rate the frequency or intensity of particular behaviors without an evaluative component.

Secondly, although all the dimensions measured in Wave Performance 360 are demonstrable behaviors, they can be displayed differently to different raters. That is, one rater may not see what another rater sees because, for example, they have not worked with the assessee on the same project or for the same period of time. This can naturally lead to differences in their ratings. Other factors including rater characteristics (e.g., leniency) and the context in which the rater operates (e.g., relationship with the assessee - boss, peer or report) may also lead to differences in ratings between raters.

14.3 Further Reference Material

Further information can be found in the 'Norms' and 'Group Trends' chapters and the Appendices of this handbook.

15.0 Validity

Wave® Performance 360 and Validity

To add to the information already provided in previous chapters, various forms of evidence are provided in this chapter to support the use of Performance 360 in evaluating work effectiveness and making valid interpretations of the scores obtained.

A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses...Ultimately, the validity of an intended interpretation...relies on all the available evidence relevant to the technical quality of a testing system.

*Standards for Educational and Psychological Testing,
 American Psychological Association 1999*

There are three sections to this chapter. The first focuses on the validity of the constructs measured in the Wave Performance 360 questionnaire by looking at the correlations of its scales with scales in a related measure of performance (convergent validity); the second section deals specifically with how the behavior and ability scales relate to superordinate constructs (i.e., global measures of effectiveness); finally, the last section centers on how Performance 360 relates to predictor measures (Wave Professional Styles and Aptitude Tests).

15.1 Wave Performance 360 and In-depth Performance 360 Measure

This section looks at the relationships between scales from Wave Performance 360 and those from an in-depth measure of the same components. This special version of Performance 360 individually assesses the 147 components of the Saville Consulting Wave Performance Culture Framework. The 45 items of Wave Performance 360 have direct counterparts in the in-depth measure, as do the higher order scales these items form (e.g., 12 behavior section scales). Performance 360 scales are, therefore, expected to correlate positively with the matched counterparts from the in-depth measure.

This part of the analysis is based on data from Project Epsom. Further information on the background of Project Epsom can be found in the 'Validity' chapter of the Saville Consulting Wave Professional Styles Handbook.

As part of Project Epsom, 208 participants rated themselves on both the Wave Performance 360 questionnaire and the in-depth Performance 360 measure. Participants were also asked to nominate two external raters to rate their work performance using the same questionnaires. The in-depth questionnaire used the same seven-point normative effectiveness rating scale as Performance 360. Ratings from the participants and their two raters were aggregated. Table 15.1 shows the correlations between Performance 360 scales and their matched counterparts from the in-depth measure. The average correlation between the two performance instruments was .65, ranging from .52 (Interpreting Data) to .82 (Working with Equipment).

Additional information on the highest other (non-matched) correlation for each Performance 360 scale is also given in Table 15.1. Of the 45 scales in Performance 360, 41 correlated highest with their own matched scale in the in-depth measure. Of the remaining four scales, two correlated equally highly with another conceptually related scale (*Working with Details* correlated with *Checking Things* ($r=.62$), and *Working with Designs* correlated with *Working with Equipment* ($r=.69$)), and two correlated slightly higher with another scale than their own counterparts (*Documenting Facts* correlated more strongly with *Working with Words* ($r=.61$) than its own counterpart ($r=.60$), and *Working with Systems* correlated more strongly with *Working with Numbers* ($r=.65$) than its own counterpart ($r=.64$)). Although *Documenting Facts* correlated highest with *Working with Words* in the in-depth measure, the next highest correlation was with its own counterpart.

The correlations between Performance 360 scales and those from the in-depth measure lend support for the construct separation of the 39 scales measured in the Wave Performance 360 questionnaire version without the ability scales, which covers only the areas of Behavior (36 dimensions) and Global (three sections). This provides evidence that raters as a group are able to separate out these different 39 components. When using the version of Wave Performance 360 with Ability Profile, there is evidence that the ability dimensions will correlate with some of the behavior dimensions (and each other), and it may be more difficult to separate these constructs so clearly. It is also important to note that this method of analysis may slightly overestimate the degree of overlap as it searches for the highest correlation, so it may find a correlation which is higher than would be expected in the population as a statistical methodological artifact.

Table 15.1 Correlations between Wave Performance 360 and In-depth Performance 360 Measure (N=208 sets of matched Performance 360 and In-depth Performance 360 assessments)

Scale	Performance 360		In-depth Performance 360		r	Highest Other Correlation	
	Mean	SD	Mean	SD		Scale	r
Examining Information	16.73	1.99	48.74	4.53	.55**	Providing Insights	.47**
Documenting Facts	16.66	2.36	49.36	5.28	.60**	Working with Words	.61**
Interpreting Data	15.89	2.09	45.54	5.21	.52**	Working with Numbers	.50**
Developing Expertise	16.77	2.19	48.85	5.42	.64**	Applying Specialist Expertise	.45**
Adopting Practical Approaches	17.56	1.86	51.50	5.29	.59**	Applying Specialist Expertise	.44**
Providing Insights	16.38	2.04	47.76	4.53	.55**	Accomplishing Objectives	.50**
Generating Ideas	15.33	2.34	42.90	5.82	.63**	Taking Action	.47**
Exploring Possibilities	15.07	2.09	44.19	5.43	.52**	Inviting Feedback	.47**
Developing Strategies	14.78	2.32	43.77	5.80	.62**	Generating Ideas	.49**
Interacting with People	17.15	2.50	47.49	5.78	.69**	Establishing Rapport	.65**
Establishing Rapport	17.83	2.39	50.46	6.44	.70**	Interacting with People	.57**
Impressing People	15.35	2.74	42.87	6.92	.62**	Conveying Self-Confidence	.50**
Convincing People	15.61	2.16	45.26	5.33	.58**	Interacting with People	.40**
Articulating Information	15.13	2.92	45.17	5.94	.68**	Directing People	.49**
Challenging Ideas	15.69	2.49	45.20	5.76	.54**	Exploring Possibilities	.46**
Making Decisions	16.19	2.06	49.16	5.21	.58**	Directing People	.48**
Directing People	15.29	2.71	45.13	6.07	.69**	Accomplishing Objectives	.49**
Empowering Individuals	16.08	2.58	47.41	5.82	.67**	Establishing Rapport ¹	.54**
Conveying Self-Confidence	15.21	2.51	45.64	5.98	.67**	Articulating Information	.54**
Showing Composure	15.78	2.80	45.69	6.95	.75**	Thinking Positively	.46**
Resolving Conflict	15.81	2.57	46.26	6.55	.72**	Establishing Rapport	.59**
Thinking Positively	16.28	2.44	48.57	5.62	.65**	Showing Composure	.47**
Embracing Change	15.94	2.19	46.05	5.71	.56**	Showing Composure	.53**
Inviting Feedback	15.13	2.32	43.69	5.09	.54**	Thinking Positively	.40**
Understanding People	17.19	2.40	49.07	5.93	.68**	Establishing Rapport ²	.60**
Team Working	17.03	2.13	48.94	5.45	.61**	Empowering Individuals	.47**
Valuing Individuals	17.03	2.28	47.77	6.23	.69**	Understanding People	.64**
Meeting Timescales	17.22	2.38	50.86	6.00	.74**	Managing Tasks	.54**
Checking Things	17.19	2.05	50.61	5.42	.72**	Meeting Timescales	.46**
Following Procedures	16.76	2.28	47.90	5.81	.69**	Checking Things	.50**
Managing Tasks	16.98	2.01	49.02	5.37	.57**	Meeting Timescales	.55**
Upholding Standards	18.08	1.82	53.15	5.54	.74**	Checking Things	.49**
Producing Output	17.64	2.09	50.56	5.10	.64**	Meeting Timescales	.58**
Taking Action	16.59	2.12	50.08	5.00	.59**	Making Decisions	.57**
Seizing Opportunities	14.28	2.79	43.14	7.60	.76**	Developing Strategies	.52**
Pursuing Goals	16.54	2.21	50.75	5.26	.62**	Providing Insights	.51**
Working with Words	16.35	2.44	81.76	10.35	.75**	Documenting Facts	.65**
Working with Numbers	14.61	2.78	72.96	10.38	.76**	Working with Systems	.60**
Working with Details	15.67	2.65	76.53	8.76	.62**	Checking Things	.62**
Working with Systems	14.62	2.52	73.83	9.46	.64**	Working with Numbers	.65**
Working with Designs	13.29	2.77	65.74	12.11	.69**	Working with Equipment	.69**
Working with Equipment	13.37	3.28	63.50	13.30	.82**	Working with Designs	.69**
Applying Specialist Expertise	16.57	1.82	51.20	4.95	.64**	Developing Expertise	.48**
Accomplishing Objectives	17.22	1.77	50.86	4.69	.63**	Taking Action	.52**
Demonstrating Potential	15.84	2.75	48.07	7.07	.75**	Developing Strategies	.49**
Mean	16.08	2.36	50.96	6.36	.65		
Median	16.19	2.34	48.74	5.78	.64		
Min	13.29	1.77	42.87	4.53	.52		
Max	18.08	3.28	81.76	13.30	.82		

** All Correlations are significant at the 0.01 level (2-tailed).
¹ Resolving Conflict also correlated at .54**.
² Resolving Conflict also correlated at .60**.

15.2 Wave Performance 360 and Superordinate Constructs

It is sometimes thought that the criterion-related validity of 360 and other such tools would be difficult to ascertain because they are in themselves criterion measures. What, then, is the criterion to assess the criteria against? Despite this circularity, important questions still need to be answered in relation to the interpretation of assessment scores. When an individual is seen as effective in certain work behaviors, what does this mean? How do these observations relate to the individual's overall job performance? Are we confident in saying that the individual is doing a good job?

In this section, we are specifically interested in how Wave Performance 360 scales relate to overall performance. This is the nearest corollary we have to criterion-related validity and with it we can address some important issues in relation to the validity of Performance 360.

Performance 360 was researched, designed and developed to measure work effectiveness, with each item being expected to underpin overall work performance. As a result, it is anticipated that there will be positive relationships between both the behavior and ability component areas and overall (global) work effectiveness.

As mentioned in the Introduction chapter of this handbook, 'Global' in the Wave Performance Culture Framework describes the broad overall effectiveness characteristics of performance at work.

15.2.1 Overall Measures of Behavior, Ability and Global Effectiveness

In the first analysis, we look at the correlations between both behavior and ability ratings given by the external rater categories (Boss, Peer and Report) and global ratings of effectiveness at work.

Three different overall scores were created: Overall Behavior, Overall Ability, and Overall Global. If the components of behavior and ability underpin overall work effectiveness as expected in Performance 360, then the Overall Behavior and the Overall Ability scores should correlate appreciably with the Overall Global score.

The Overall Behavior score was created by first averaging the ratings for all raters within one category to create one score for each of the 36 behavior dimensions. These 36 behavior dimension scores were then summed to create an Overall Behavior score. Similarly, an average score on an ability dimension was obtained from the ratings of all raters in one category. These six average scores were added together to create the Overall Ability score. Global Performance was aggregated from the scores of the three global items in the same manner. These three overall scores were calculated for each of the three rater categories, Boss, Peer and Report.

Table 15.2 shows the correlation matrix between the Overall Behavior and Global Performance scores from the Boss, Peer and Report categories. The analyses included only Performance 360 assessment sets that had at least 1 boss, 2 peers and 2 reports, and 1,320 assessment sets met this criterion (total number of participants N=14,355).

The matrix shows strong correlations between the Overall Behavior score and Global Performance within the rater categories, namely .77 for boss, .81 for peer and .86 for report. Additionally, positive correlations of moderate size were found between categories. For example, the sum of all the behaviors as rated by a peer correlated .35 with the report's view of their overall effectiveness.

These correlations indicate that those rated higher across a greater number of behavior dimensions by a particular group of raters tend also to be rated higher at Overall (Global) Effectiveness by other rater groups. This lends support to the behaviors of Performance 360 underpinning overall effectiveness at work.

Table 15.2 Overall Behavior Score against Overall Global Performance Score for Three Rater Categories (N=14,355)

Overall Behavior	Overall Global Performance		
	Boss	Peer	Report
Boss	.77**	.37**	.31**
Peer	.31**	.81**	.35**
Report	.23**	.31**	.86**

** Correlation is significant at the 0.01 level (two-tailed).

Out of the 1,320 assessments, 385 assessments included the Ability Profile (total number of participants N=3,849). Based on data from the 385 assessments, Table 15.3 shows the correlation coefficients between Overall Ability and Overall (Global) Performance scores from Boss, Peer and Report categories. The correlation coefficients within the same rater categories were .37 for boss, .42 for peer and .47 for report respectively.

Table 15.3 Overall Ability Score against Overall Global Performance Score for Three Rater Categories (N=3,849)

Overall Ability	Overall Global Performance		
	Boss	Peer	Report
Boss	.37**	.04	.01
Peer	.09	.42**	.08
Report	.10	.12*	.47**

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

The link between Overall Ability scores and Overall Global Performance is less clear. We know that cognitive ability tests are strong predictors of overall performance, so this weaker link may appear counter-intuitive. However, given the lower reliability of these areas (cf. 'Reliability' chapter), it would seem that these attributes may be more difficult to assess using subjective rating measures.

If a detailed understanding of an individual's aptitude is required, we would recommend

that measures of cognitive ability (i.e., Saville Consulting Aptitude Tests) should be deployed in conjunction with Wave Performance 360. Nevertheless, the ability areas do show a degree of correlation with Overall Global Performance, and it may be useful for the individual to reflect on the ratings they are given on the ability areas by others to understand the impression they are creating.

The correlation between the Performance 360 Ability areas and tests is further discussed in section 15.3.2 of this chapter.

15.2.2 Individual Components of Behavior, Ability and Overall Global Effectiveness

In the second analysis, we consider how individual behavior and ability dimensions of Performance 360 relate to overall or global effectiveness, based on the sample used in the first analysis. Again, this is done for the different rater groups.

Table 15.4 shows the correlations between behavior and ability ratings at dimension level and Global Performance ratings for the Boss, Peer and Report categories. The first three columns display the correlations for Boss behavior and ability ratings with Global Performance ratings from each rater category, the second group of three columns displays the correlations for Peers, and the last three columns for Reports.

The average correlation of behavior and ability with Global Performance for each category correlation coefficient was .43 for Boss, .51 for Peer and .60 for Report (columns shown with dark blue background). Pursuing Goals was the dimension that had highest within rater category correlations with Global Performance for all three rater categories (.64 for Boss, .72 for Peers and .78 for Reports).

As in the previous analysis of overall scores, there were positive correlations between the rater categories. These positive correlations provide support for the componential model of performance, with the behavior and ability components correlating positively with Global Performance, suggesting that these behavioral components underpin overall performance at work in line with Performance 360's development aims.

There are two exceptions to this pattern of positive correlations with two of the ability dimensions: Working with Designs, and Working with Equipment. Both have weak correlations (approximating zero) with Global Performance ratings (whether they were within or between rater categories) and two correlations were statistically significant in the negative direction.

It is possible that these two dimensions were not as relevant as other ability dimensions to the majority of job roles of the individuals being rated in this analysis. For the minority of clients who are using the option of Performance with the Ability Profile, it may be worth considering whether these two ability areas hold particular relevance to the job in question, as they are unlikely to underpin Overall (Global) Performance in many roles.

Table 15.4 Behavior and Ability Dimension Ratings from Boss, Peer and Report categories against their ratings on Global Performance (N=14,355; 3,849^a)

Dimension	Boss Dimension Rating against...			Peer Dimension Rating against...			Report Dimension Rating against...		
	Boss Global Rating	Peer Global Rating	Report Global Rating	Boss Global Rating	Peer Global Rating	Report Global Rating	Boss Global Rating	Peer Global Rating	Report Global Rating
Examining Information	.51**	.27**	.23**	.28**	.59**	.29**	.23**	.25**	.65**
Documenting Facts	.44**	.24**	.19**	.22**	.54**	.23**	.20**	.22**	.62**
Interpreting Data	.50**	.24**	.16**	.24**	.55**	.21**	.19**	.23**	.64**
Developing Expertise	.48**	.27**	.19**	.22**	.58**	.20**	.23**	.28**	.67**
Adopting Practical Approaches	.47**	.19**	.16**	.20**	.58**	.26**	.18**	.23**	.68**
Providing Insights	.56**	.28**	.21**	.28**	.67**	.31**	.21**	.27**	.74**
Generating Ideas	.48**	.28**	.18**	.20**	.57**	.22**	.20**	.27**	.67**
Exploring Possibilities	.53**	.31**	.20**	.22**	.61**	.23**	.20**	.27**	.70**
Developing Strategies	.56**	.30**	.23**	.25**	.62**	.22**	.22**	.28**	.68**
Interacting with People	.40**	.21**	.20**	.15**	.46**	.21**	.16**	.23**	.60**
Establishing Rapport	.28**	.17**	.19**	.07*	.34**	.17**	.09**	.16**	.52**
Impressing People	.43**	.25**	.19**	.22**	.50**	.19**	.17**	.26**	.57**
Convincing People	.48**	.26**	.25**	.20**	.59**	.26**	.17**	.24**	.69**
Articulating Information	.46**	.23**	.18**	.24**	.55**	.26**	.18**	.28**	.67**
Challenging Ideas	.50**	.22**	.16**	.24**	.59**	.25**	.19**	.23**	.63**
Making Decisions	.54**	.23**	.22**	.27**	.63**	.29**	.21**	.29**	.71**
Directing People	.52**	.23**	.28**	.26**	.60**	.31**	.20**	.28**	.71**
Empowering Individuals	.46**	.23**	.27**	.18**	.54**	.28**	.15**	.25**	.68**
Conveying Self-Confidence	.43**	.23**	.17**	.26**	.54**	.22**	.19**	.25**	.62**
Showing Composure	.30**	.14**	.11**	.17**	.40**	.16**	.11**	.16**	.49**
Resolving Conflict	.36**	.15**	.15**	.10**	.40**	.18**	.09**	.13**	.55**
Thinking Positively	.38**	.19**	.15**	.17**	.45**	.19**	.16**	.19**	.60**
Embracing Change	.50**	.22**	.16**	.25**	.58**	.22**	.18**	.23**	.69**
Inviting Feedback	.42**	.17**	.12**	.11**	.49**	.20**	.11**	.18**	.61**
Understanding People	.33**	.16**	.16**	.08**	.39**	.20**	.09**	.17**	.58**
Team Working	.45**	.19**	.20**	.12**	.50**	.24**	.11**	.22**	.65**
Valuing Individuals	.33**	.12**	.10**	.04	.41**	.18**	.10**	.17**	.59**
Meeting Timescales	.45**	.26**	.20**	.27**	.54**	.25**	.20**	.25**	.57**
Checking Things	.46**	.23**	.18**	.28**	.55**	.21**	.20**	.21**	.60**
Following Procedures	.40**	.14**	.11**	.22**	.47**	.17**	.13**	.16**	.55**
Managing Tasks	.50**	.21**	.18**	.31**	.60**	.26**	.20**	.25**	.65**
Upholding Standards	.45**	.19**	.20**	.20**	.54**	.22**	.16**	.24**	.63**
Producing Output	.54**	.26**	.21**	.33**	.66**	.32**	.24**	.28**	.72**
Taking Action	.62**	.31**	.24**	.31**	.69**	.31**	.26**	.32**	.76**
Seizing Opportunities	.46**	.19**	.16**	.15**	.51**	.18**	.18**	.20**	.61**
Pursuing Goals	.64**	.30**	.26**	.31**	.72**	.30**	.24**	.30**	.78**
Working with Words ^a	.39**	.16**	.15**	.13*	.41**	.12*	.15**	.21**	.51**
Working with Numbers ^a	.30**	.07	.08	.14**	.37**	.13*	.05	.08	.35**
Working with Details ^a	.30**	.05	.03	.11*	.34**	.04	.08	.05	.35**
Working with Systems ^a	.26**	.04	.01	.10	.35**	.07	.11*	.09	.37**
Working with Designs ^a	.15**	-.07	-.12*	.01	.20**	.02	.01	.03	.23**
Working with Equipment ^a	.12*	-.06	-.10	-.11*	.13*	-.04	.01	.04	.22**
Mean	.43	.20	.16	.19	.51	.21	.16	.21	.60
Median	.46	.22	.18	.21	.54	.22	.18	.23	.63
Min	.12	-.07	-.12	-.11	.13	-.04	.01	.03	.22
Max	.64	.31	.28	.33	.72	.32	.26	.32	.78

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

15.2.3 Further Breakdown of the Components of Global Effectiveness

Having established the relationships with Global Performance, we now consider in more detail the correlations of each behavior and ability dimension with the three different dimensions that comprise Global Performance in the Wave Performance Culture Framework: Applying Specialist Expertise, Accomplishing Objectives and Demonstrating Potential.

The aim of studying the relationships with the three global sections was to explore whether a behavior or ability dimension would correlate strongly with one or more individual aspects of Global Performance.

There were two categories of correlation calculated for the purpose of these analyses. The within rater category is based on correlations of ratings given by the same raters. So, for example, the first correlation ($r=.52$) in Table 15.5 is the average of all the correlations between Examining Information and Applying Specialist Expertise by the same raters within each rater category. The between rater category then is the average correlation between different raters (i.e., the average ratings for different rater categories). Referring back to the example provided above, we note that for Examining Information and Applying Specialist Expertise the average correlation between the averages in different rater categories is $r=.22$.

For readers interested in a breakdown of how each rater category fared individually, the three separate matrices can be found in the Appendices.

The overall pattern of positive correlations between the individual behavior and ability components further reinforces the componential model of Performance 360 which is designed to underpin Overall (Global) Effectiveness at work.

Looking at Table 15.5, it is clear that some behavior dimensions correlated highly with all the three global sections. An example is Developing Expertise. It would be natural to assume a link between Developing Expertise and the global effectiveness area of Applying Specialist Expertise, but in fact the data also indicates that Developing Expertise is likely to have an important role to play in both Accomplishing Objectives and Demonstrating Potential. In other words, getting the job done and getting ahead at work is likely to require learning and the development of expertise (it is not just something which is good for those who are considered experts!).

Other behavior dimensions show a more differentiated pattern of correlations across the three different global sections. For example, Impressing People correlated less strongly with the global dimensions of Applying Specialist Expertise than with Demonstrating Potential.

The correlations between rater categories provide additional support for the behavior and ability dimensions that underpin overall effectiveness, as the analysis is based on independent groups of raters. It can be seen from the table that the top five dimensions that underpin Applying Specialist Expertise are Examining Information, Upholding Standards, Documenting Facts, Interpreting Data and Exploring Possibilities. The top six dimensions that underpin Accomplishing Objectives are Taking Action, Pursuing Goals, Producing Output, Meeting Timescales, Managing Tasks (=5th) and Providing Insights (=5th), and the top six for Demonstrating Potential are Taking Action, Pursuing Goals, Producing Output, Directing People, Making Decisions (=5th) and Impressing People (=5th).

These observations indicate that although some behavior and ability constructs underpin all three aspects of Global Performance, other constructs are more related to just one or two aspects. The general pattern of positive correlations between the behaviors and the different global effectiveness components suggests that the behavior (and ability) components generally underpin the different components of Overall (Global) Effectiveness in Performance 360.

Table 15.5 Average Correlation Coefficients (Within Rater Categories and Between Rater Categories) between Behavior and Ability Dimension Ratings and Global Dimensions (N=14,355; 3,849^a)

Dimension	Applying Specialist Expertise		Accomplishing Objectives		Demonstrating Potential	
	Within Rater Categories	Between Rater Categories	Within Rater Categories	Between Rater Categories	Within Rater Categories	Between Rater Categories
Examining Information	.52	.22	.51	.22	.44	.21
Documenting Facts	.48	.20	.48	.18	.40	.16
Interpreting Data	.52	.20	.48	.17	.42	.16
Developing Expertise	.50	.19	.48	.18	.47	.20
Adopting Practical Approaches	.47	.17	.54	.18	.44	.17
Providing Insights	.51	.19	.59	.23	.54	.23
Generating Ideas	.44	.16	.49	.18	.49	.22
Exploring Possibilities	.54	.20	.52	.20	.47	.20
Developing Strategies	.47	.18	.56	.22	.53	.23
Interacting with People	.34	.11	.45	.18	.43	.19
Establishing Rapport	.31	.10	.35	.13	.29	.12
Impressing People	.28	.09	.41	.18	.52	.24
Convincing People	.44	.15	.53	.20	.50	.22
Articulating Information	.44	.16	.47	.19	.48	.22
Challenging Ideas	.47	.17	.49	.18	.47	.18
Making Decisions	.46	.15	.58	.24	.52	.24
Directing People	.41	.14	.59	.25	.53	.25
Empowering Individuals	.44	.15	.53	.21	.44	.20
Conveying Self-Confidence	.34	.12	.44	.18	.52	.23
Showing Composure	.32	.11	.36	.12	.32	.12
Resolving Conflict	.37	.10	.40	.13	.33	.11
Thinking Positively	.33	.10	.46	.17	.40	.17
Embracing Change	.42	.12	.55	.19	.51	.20
Inviting Feedback	.41	.11	.47	.13	.40	.13
Understanding People	.40	.13	.40	.14	.30	.10
Team Working	.45	.14	.50	.17	.40	.15
Valuing Individuals	.41	.12	.41	.11	.31	.08
Meeting Timescales	.38	.15	.52	.24	.41	.20
Checking Things	.47	.19	.49	.20	.39	.16
Following Procedures	.41	.12	.46	.15	.33	.11
Managing Tasks	.44	.15	.57	.23	.46	.20
Upholding Standards	.52	.22	.50	.17	.36	.12
Producing Output	.47	.17	.60	.25	.53	.25
Taking Action	.47	.16	.65	.27	.59	.28
Seizing Opportunities	.33	.06	.48	.16	.48	.20
Pursuing Goals	.50	.17	.69	.26	.59	.27
Working with Words ^a	.47	.19	.35	.10	.31	.11
Working with Numbers ^a	.29	.07	.30	.06	.27	.10
Working with Details ^a	.29	.08	.28	.01	.27	.06
Working with Systems ^a	.30	.06	.28	.04	.25	.08
Working with Designs ^a	.16	-.05	.16	-.04	.17	.02
Working with Equipment ^a	.15	-.04	.12	-.05	.14	-.01
Mean	.41	.13	.46	.16	.41	.17
Median	.44	.15	.48	.18	.43	.19
Min	.15	-.05	.12	-.05	.14	-.01
Max	.54	.22	.69	.27	.59	.28

15.3 Correlations with Predictors

As Performance 360 is at the criterion rather than the predictor end, we have to seek methods linking it to other measures to gain an improved understanding of the measure’s validity. One such approach is to link Performance 360 to predictors which were designed to forecast the attributes it measures.

15.3.1 Correlation with Wave Professional Styles

For the purpose of this analysis, two overall scales were created from Performance 360 in the Project Epsom validation study (N=308) based on the rating from one external rater. The first overall score was the average of 36 behavior dimensions, and the second overall scale was the three items of the Global Effectiveness scale. These were correlated with an a priori equation from Wave Professional Styles based on a mapping to the Great Eight. Please refer to the ‘Validity’ Chapter of the Saville Consulting Wave Professional Styles Handbook for further information.

Table 15.6 Criterion-Related Validity of Wave Professional Styles based on A Priori Hypotheses with two Performance 360 Overall Scales (N=308)

Performance 360 Overall Scale	Criterion Related Validity (against Wave Professional Styles)
Wave Behavior Dimensions (36 Items)	.25**
Global Effectiveness (3 Items)	.32**

*** All Correlations are significant at the 0.01 level (2-tailed).
No corrections were made (e.g., criterion unreliability, restriction of range, etc.)
.32 is the published overall effectiveness figure documented in Project Epsom (MacIver et al., 2008)*

The results presented in Table 15.6 indicate a convergence between the two overall scores and Wave Professional Styles. This provides support for the three component Overall Global Performance scale which, despite being a short scale, shows a stronger correlation with Wave Professional Styles.

15.3.2 Correlation with Aptitude Measures

In previous sections of this chapter, it has been discussed that ability areas may be more difficult to assess than behaviors. This section aims to investigate further the validity of ability ratings by correlating Saville Consulting aptitude test results directly with performance ratings on matched ability criteria.

As part of Project Epsom, participants completed three aptitude tests: Swift Analysis Aptitude, Swift Comprehension Aptitude and Swift Technical Aptitude. Each of these tests contains three sub-tests. Swift Analysis Aptitude covers verbal, numerical and diagrammatic analysis; Swift Comprehension Aptitude covers verbal comprehension, numerical comprehension and error checking; and Swift Technical Aptitude comprises spatial, mechanical and diagrammatic reasoning sub-tests. The six ability areas are measured across these three combined assessments (as verbal, numerical and

diagrammatic areas are covered in two tests). Performance ratings in the six ability areas were also obtained from participants themselves and their nominated external raters using the Ability Profile in the Wave Performance 360 questionnaire.

As shown in Table 15.7, participants' total and sub-test scores for the three tests were correlated with both external and self-ratings on the six ability areas. Test scores were expected to correlate with their respective matched ability criteria, as indicated by the blue boxes in the matrix. Results show that positive correlations were found between all aptitude test scores and their respective matched ability criteria. For each aptitude test score, the average of the hypothesized correlations (i.e., the blue boxes) and the average of other correlations not hypothesized (i.e., boxes which are not highlighted) were also calculated and compared. In all cases, the average hypothesized correlations (average $r=.18$) were higher than the average of those not hypothesized (average $r=.08$). These results add support to the construct independence of these ability areas.

Although the correlations between the Diagrammatic sub-test in Swift Analysis Aptitude and Working with Systems, and between the Spatial sub-test in Swift Technical Aptitude and Working with Designs were in positive direction as hypothesized, they were less strong and not significant. This may reflect that these two ability areas are less observable and are difficult to assess.

Some differences can be seen between correlations with self-rating and external rating on the six ability criteria. Such differences once again suggest that self-ratings can be fairly different from the impressions individuals make of their ability on others. In some cases, when using aptitude test scores as benchmark, self-ratings appear to be less accurate than ratings given by others.

Table 15.7 Criterion-Related Validity of Swift Analysis Aptitude, Swift Comprehension Aptitude and Swift Technical Aptitude Scores against Matched Work Performance Criteria (N=308)

		Working with Words		Working with Numbers		Working with Details		Working with Systems		Working with Designs		Working with Equipment	
		Self	Rater	Self	Rater	Self	Rater	Self	Rater	Self	Rater	Self	Rater
Swift Analysis Aptitude	Total Score	.17**	.24**	.20**	.23**	.09	.17**	.04	.20**	-.02	.11	-.07	.08
	Verbal	.23**	.27**	.10	.22**	.06	.20**	-.02	.17**	-.02	.09	-.09	.06
	Numerical	.07	.17**	.22**	.20**	.00	.08	.06	.17**	.03	.09	-.03	.07
	Diagrammatic	.05	.07	.13*	.08	.14*	.08	.06	.10	-.06	.06	-.03	.05
Swift Comprehension Aptitude	Total Score	.21**	.24**	.20**	.22**	.16**	.26**	.05	.17**	-.04	.05	-.01	.08
	Verbal	.20**	.24**	.11	.24**	.09	.24**	.01	.18**	-.05	.05	-.06	.05
	Numerical	.14*	.16**	.23**	.18**	.15**	.18**	.11*	.11	.03	.08	.04	.05
	Error Checking	.15**	.16*	.14*	.09	.12*	.18**	.00	.11*	-.07	-.01	.00	.08
Swift Technical Aptitude	Total Score	.11	.10	.18**	.17**	.08	.11*	.12*	.22**	.13*	.19**	.13*	.14*
	Spatial	.08	.05	.04	.08	.06	.03	.05	.10	.10	.06	.09	.03
	Mechanical	.00	.03	.18**	.15**	.06	.09	.16**	.16**	.18**	.24**	.23**	.24**
	Diagrammatic	.14*	.12*	.13*	.13*	.04	.11	.04	.19**	.00	.10	-.03	.02

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

Overall, these results show that, although some ability areas may be more difficult to appraise, performance ratings in ability areas are supported by objective measures of cognitive ability. Therefore, to reiterate, measures of cognitive ability should be used together with Wave Performance 360 if a better understanding of an individual's aptitude is required. As the ratings of others on ability are more likely to be accurate than self-perceptions, it may be useful for an individual to understand the perceptions of their abilities provided by others.

15.4 Validity Summary

This chapter has introduced different strands of validity evidence for Wave Performance 360, providing support for the use of this instrument in the evaluation of work effectiveness by external raters.

The chapter first focused on the constructs measured in Wave Performance 360, in terms of how they relate to an in-depth Performance 360 measure. The bivariate correlations indicated that Wave Performance 360 aligns well to the in-depth measure for scales (average $r=.65$). For the more rarely used form of Performance 360 with the Ability Profile, 41 out of 45 scales in Wave Performance 360 correlated highest with their own counterpart construct in the in-depth measure. For Performance 360 without the Ability Profile, evidence has been provided to support construct independence of all 39 Behavior dimensions and Global sections.

The second section of this chapter presented how Wave Performance 360 aligns with a superordinate construct, Global Performance. Bivariate correlations showed positive relationships between Overall Behavior ratings and Global Performance ratings given by different rater categories (Boss, Peer and Report). The correlations were strongest when the two overall ratings came from within the same rater categories (average $r=.81$). The correlations between rater categories were also positive (average $r=.31$). Similar trends were found at behavior dimension level. For ability ratings, both within-category correlations (average $r=.42$) and between-category correlations (average $r=.07$) between Overall Ability Rating and Global Performance were weaker than those for Overall Behavior Rating, and similar trends were found at ability dimension level. In addition, positive correlations for behavior and ability scales were found when breaking down Global Performance into three sub-scales. All the results from this section support the notion that effectiveness of the different behavior and ability constructs on Performance 360 underpins overall or global effectiveness.

The final section presents information which supports the Global Performance construct through the correlation of the Performance 360 measures with Wave Professional Styles and Saville Consulting aptitude tests. Further correlations with other personality measures are detailed in the Wave Professional Styles Handbook.

In conclusion, this chapter has presented evidence for the validity of Wave Performance 360. The instrument has been shown to measure a range of behavior and ability constructs that are robust yet independent of each other. It is argued that the association between the behavior and ability constructs and global effectiveness makes for an evaluation of work performance which is designed to reflect an individual's overall job performance as well as the components which contribute to their effectiveness.

16.0 Group Trends

This chapter is devoted to presenting group trends on Saville Consulting Wave® Performance 360, including age, gender and ethnicity. Trends are presented for average (mean) scores and internal consistency reliabilities for different groups.

Average Scores

The information presented is from actual usage data of Wave Performance 360, and as a result the differences may reflect differences in composition of the different groups on other variables. For example, age differences may be related to maturational effects or could be related to longer tenure in organizations. They could also result from generational differences and differences in the composition of other variables, including gender and job type. Similarly, gender and ethnic differences could reflect other biographical differences in the composition of these groups.

When looking at differences it is useful to remind ourselves of the scale of any differences, and the impact, if any, these should have on profile interpretation.

In the following graphs and supporting text, Cohen's *d* is referred to where .20 of an SD (.40 of a sten) is a small effect, .50 of an SD (1 sten) is a medium or moderate effect size and .80 of an SD (1.60 stens) is a large effect size.

Internal Consistency

Alternate form usually provides a more appropriate and realistic estimate of questionnaires' reliability than internal consistency. Unfortunately, in actual usage data, it is the exception rather than the rule, so internal consistency is used here.

16.1 Methodology

This section outlines the methodology in creating the different age, gender and ethnicity groups.

For age trends, participants across all rater categories were split into two age groups. The first group consisted of participants who were aged under 40 years at the time of completing the assessment ('Under 40'), and the second group consisted of participants who were 40 years of age or over at the time of completion ('40 or Over').

As Performance 360 is a multi-source assessment, group trends are considered in terms of different rater categories, i.e., rater (including Boss, Peer, Report and Other categories) or assessee (Self). In this section, age trends will be looked at from three different perspectives: (1) Rater trend - comparison between the ratings 'Under 40' and '40 or Over' raters give to others (Figures 16.1, 16.2 and 16.3); (2) Assessee trend - comparison between the ratings individuals 'Under 40' and '40 or Over' are given by others (Figures 16.4 and 16.5); and (3) comparison between the ratings 'Under 40' and '40 or Over' raters give 'Under 40' and '40 or Over' assessees (Figures 16.6, 16.7, 16.8 and 16.9).

As with age trends, gender trends can be considered from three different perspectives: (1) Rater trend - comparison between male and female raters (Figures 16.10, 16.11 and 16.12); (2) Assessee trend - comparison between the ratings male and female assesseees are given by others (Figures 16.13 and 16.14); and (3) comparison between the ratings male and female raters give male and female assesseees (Figures 16.15, 16.16, 16.17 and 16.18).

In order to look at group trends in terms of ethnicity, participants across all rater categories were split into three ethnic groups: UK White, Other White and of Other Ethnicities. UK White included individuals who reported themselves as White British and White English. Other White included individuals who reported themselves as White, White Australian, White Canadian, White European etc., and those of Other Ethnicities included individuals from other cultural backgrounds and those of mixed backgrounds.

Ethnicity group trends are considered from two different perspectives: (1) Rater trend - comparison between raters of different ethnic groups (Figures 16.19, 16.20 and 16.21); and (2) Assessee trend - comparison between the ratings individuals of different ethnic groups are given (Figures 16.22 and 16.23). Due to unequal sample sizes, no comparison is made between the ratings raters of different ethnic groups give assesseees of different ethnic groups.

16.2 Age Trends

Figure 16.1 Mean Raw Scores given by Raters (i.e., not Self) who are 'Under 40' versus '40 or Over'

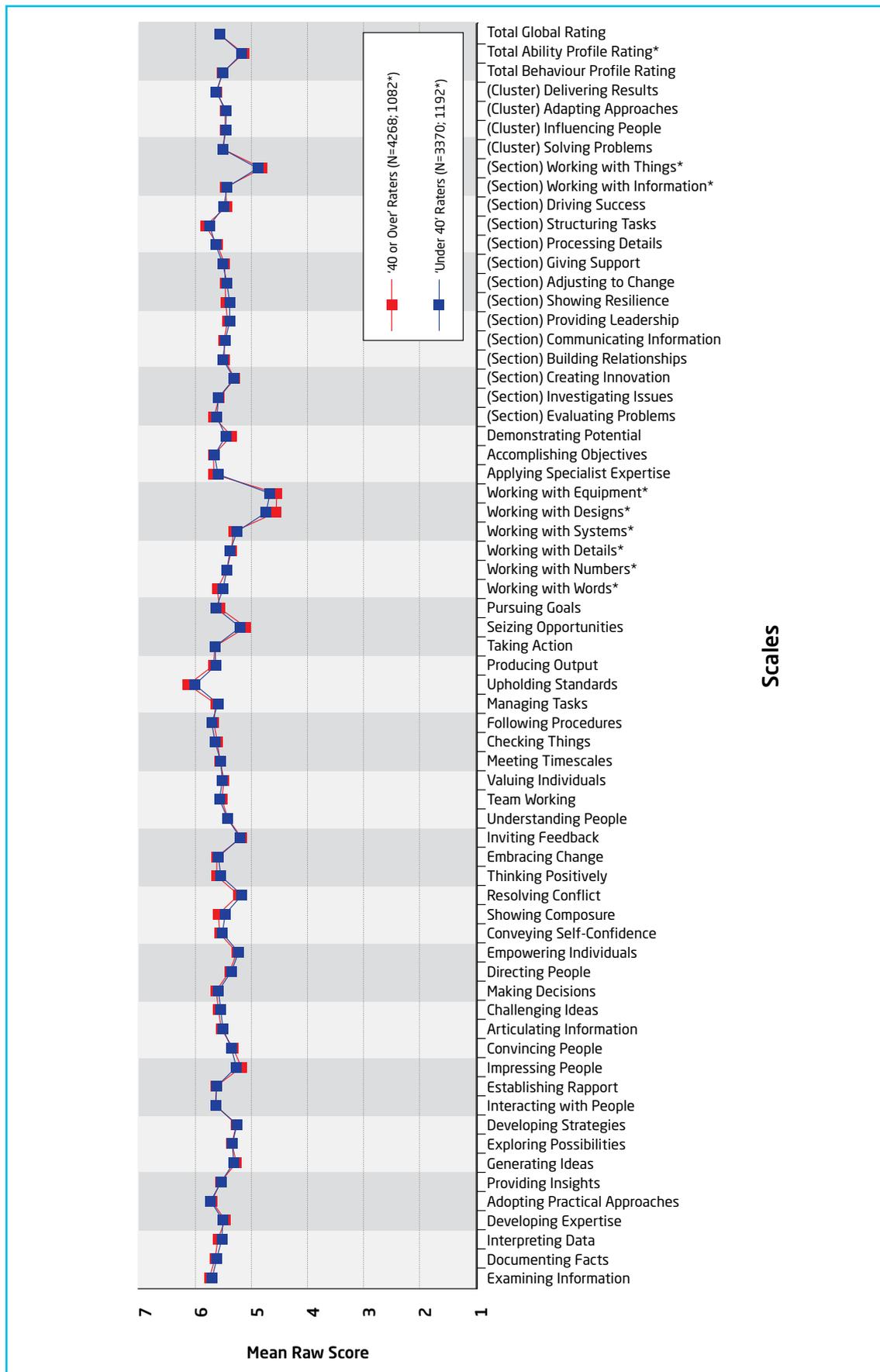


Figure 16.2 Mean Sten Scores given by Raters (i.e., not Self) who are 'Under 40' versus '40 or Over'

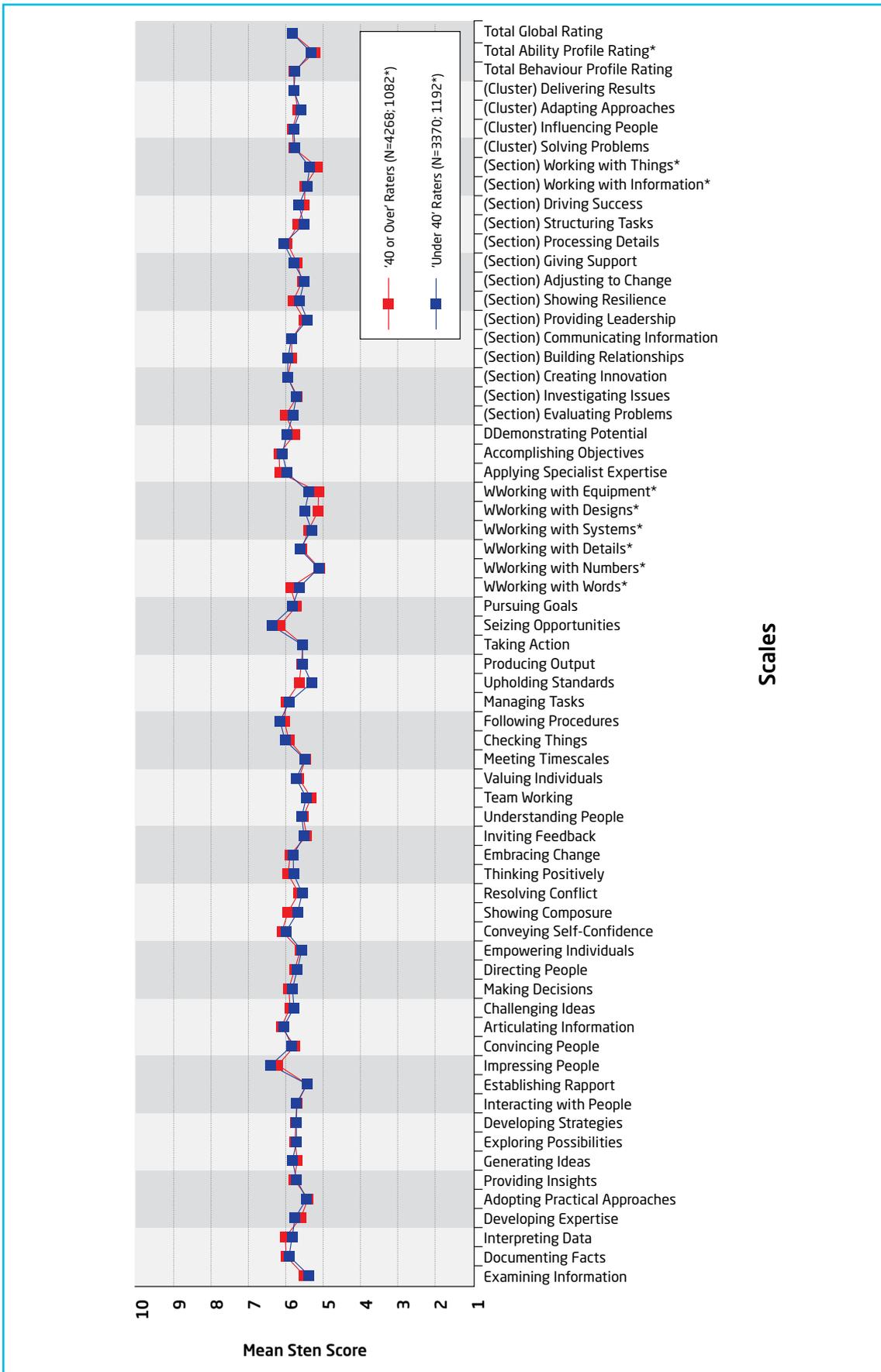
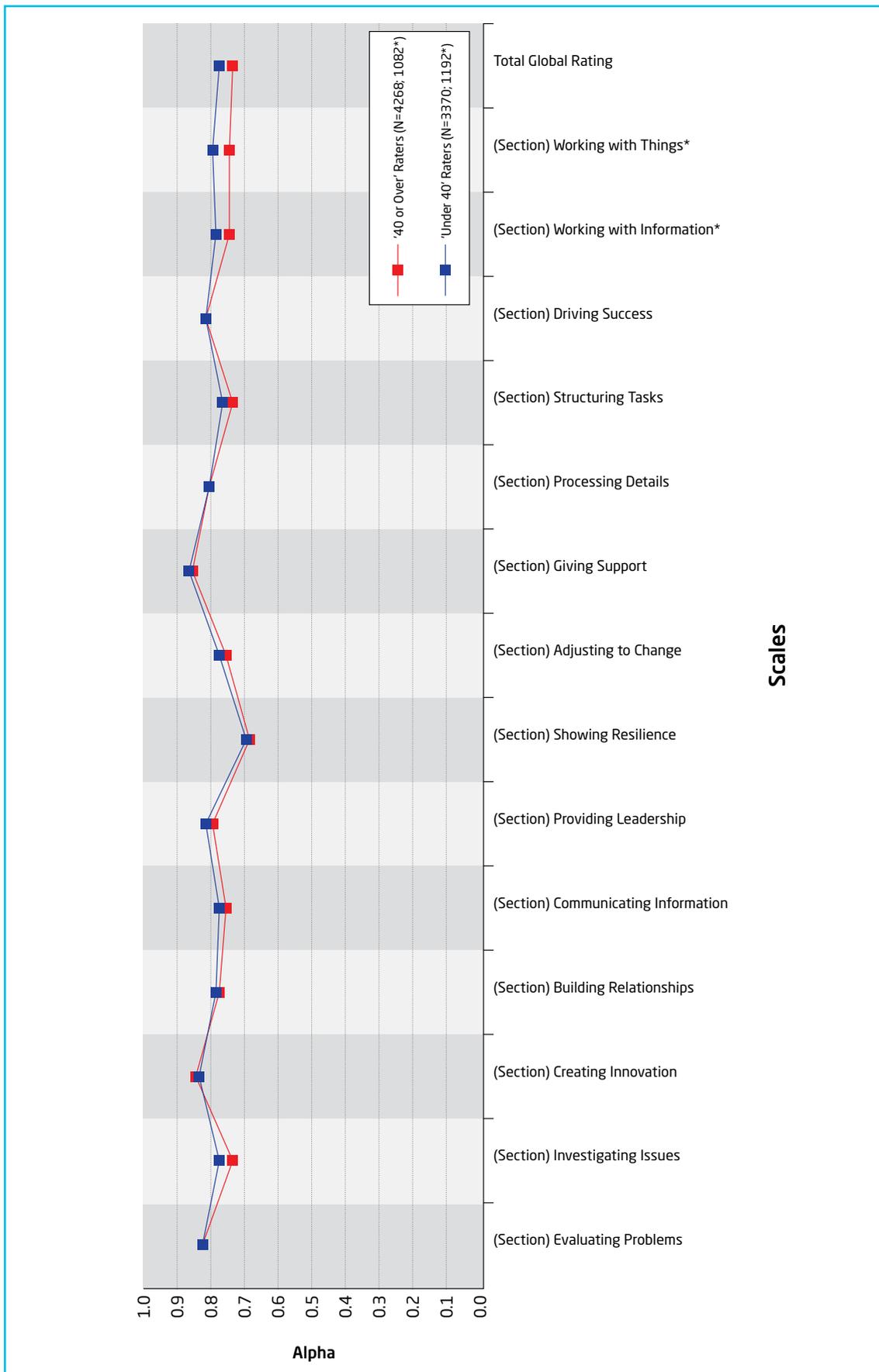
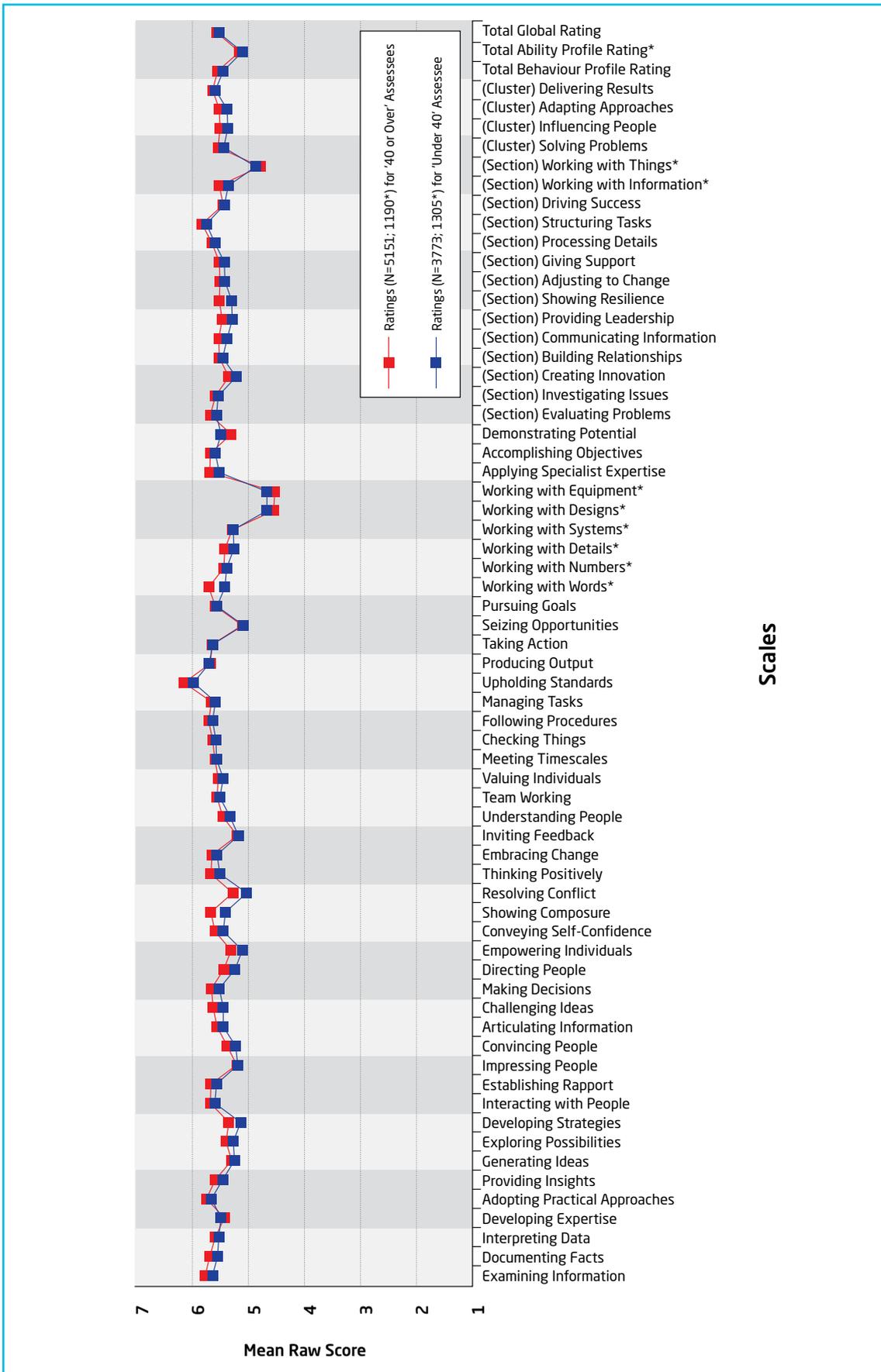


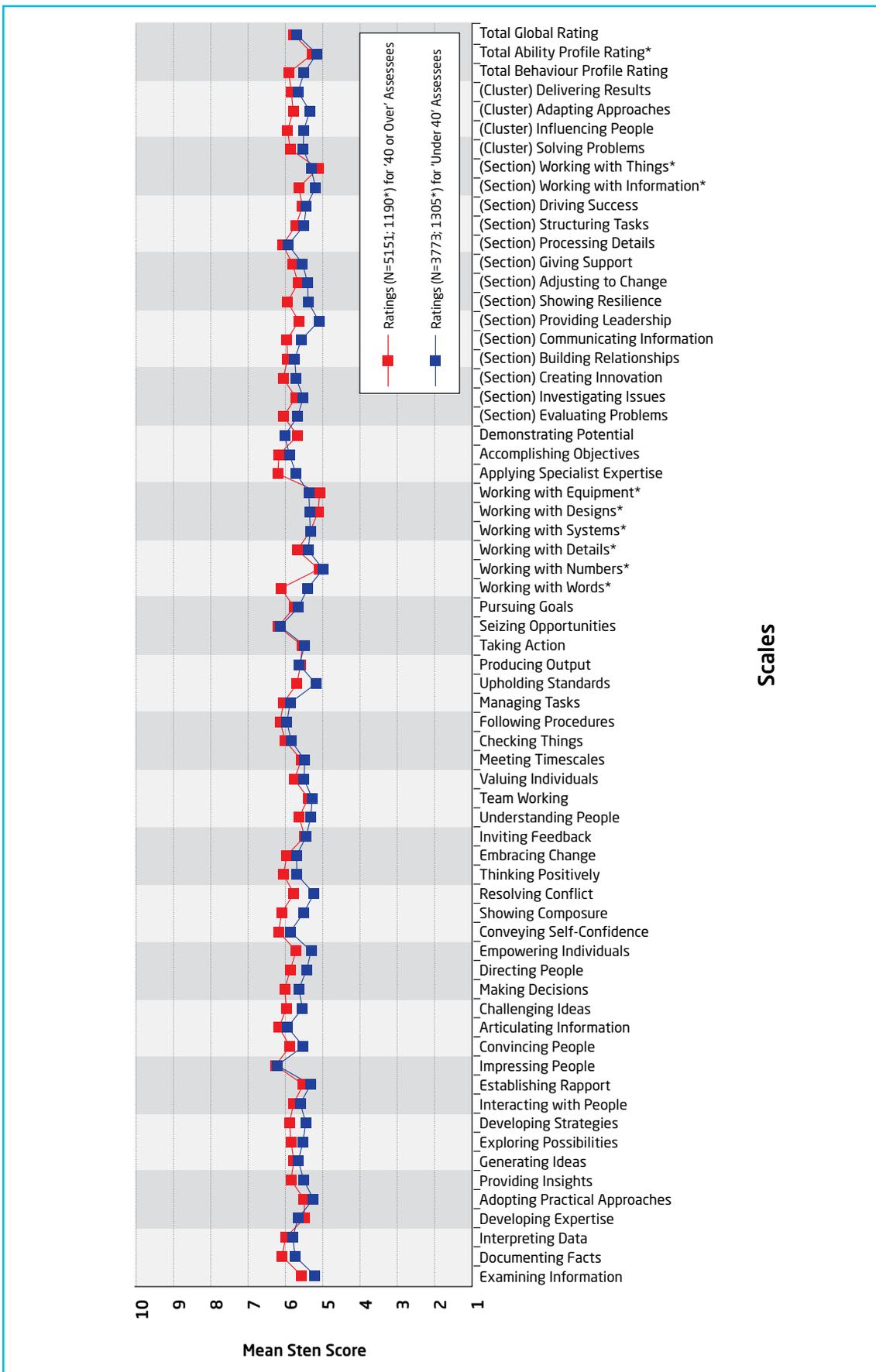
Figure 16.3 Alpha (Internal Consistency Reliabilities) for the Ratings given by Raters who are 'Under 40' versus '40 or Over'



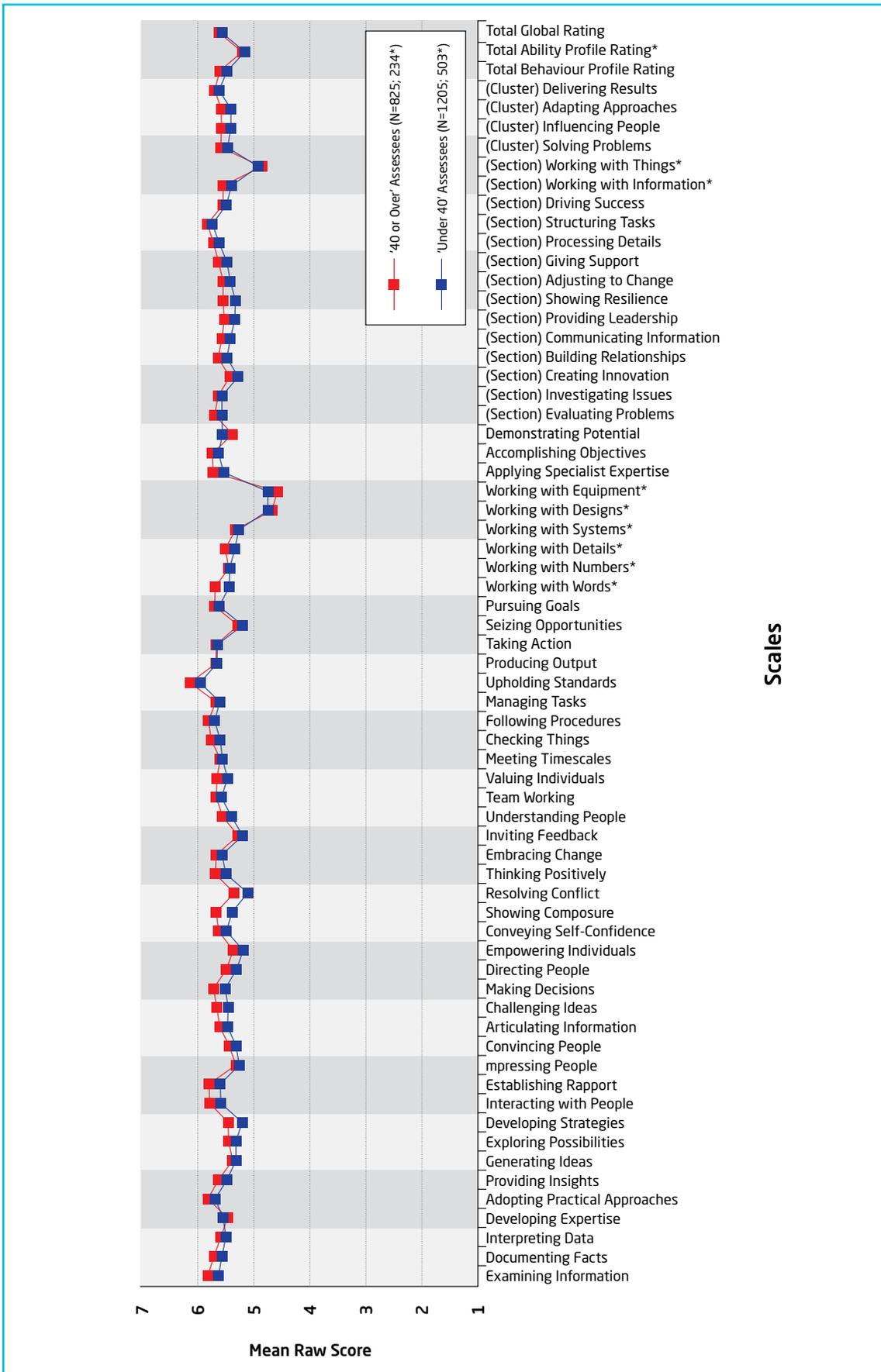
Figures 16.4 Mean Raw Scores given by Raters to Assesseees who are 'Under 40' versus '40 or Over'



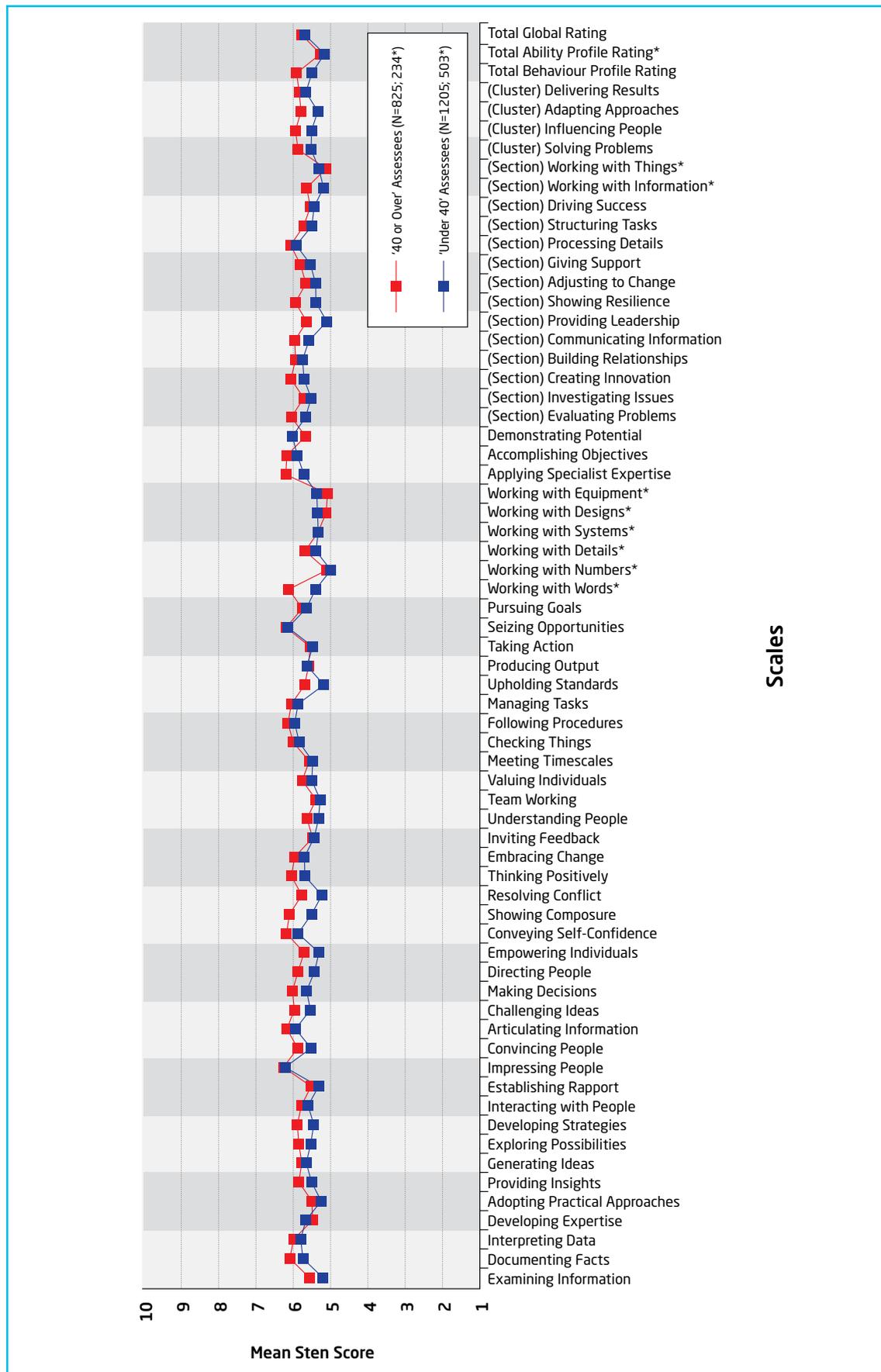
Figures 16.5 Mean Sten Scores given by Raters to Assesseees who are 'Under 40' versus '40 or Over'



Figures 16.6 Mean Raw Scores given by Raters aged Under 40 to Assesseees who are 'Under 40' versus '40 or Over'

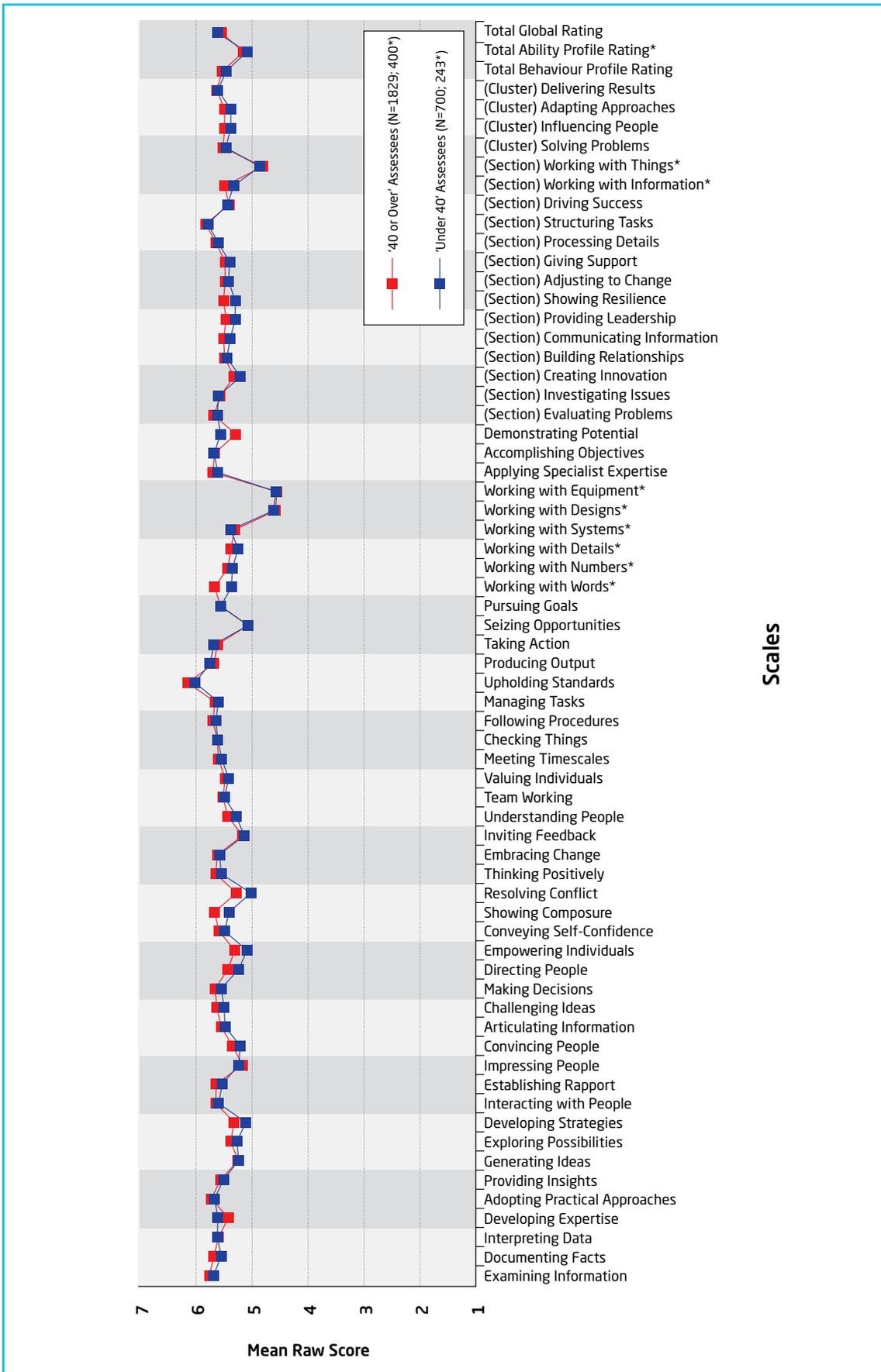


Figures 16.7 Mean Sten Scores given by Raters aged Under 40 to Assesseees who are 'Under 40' versus '40 or Over'

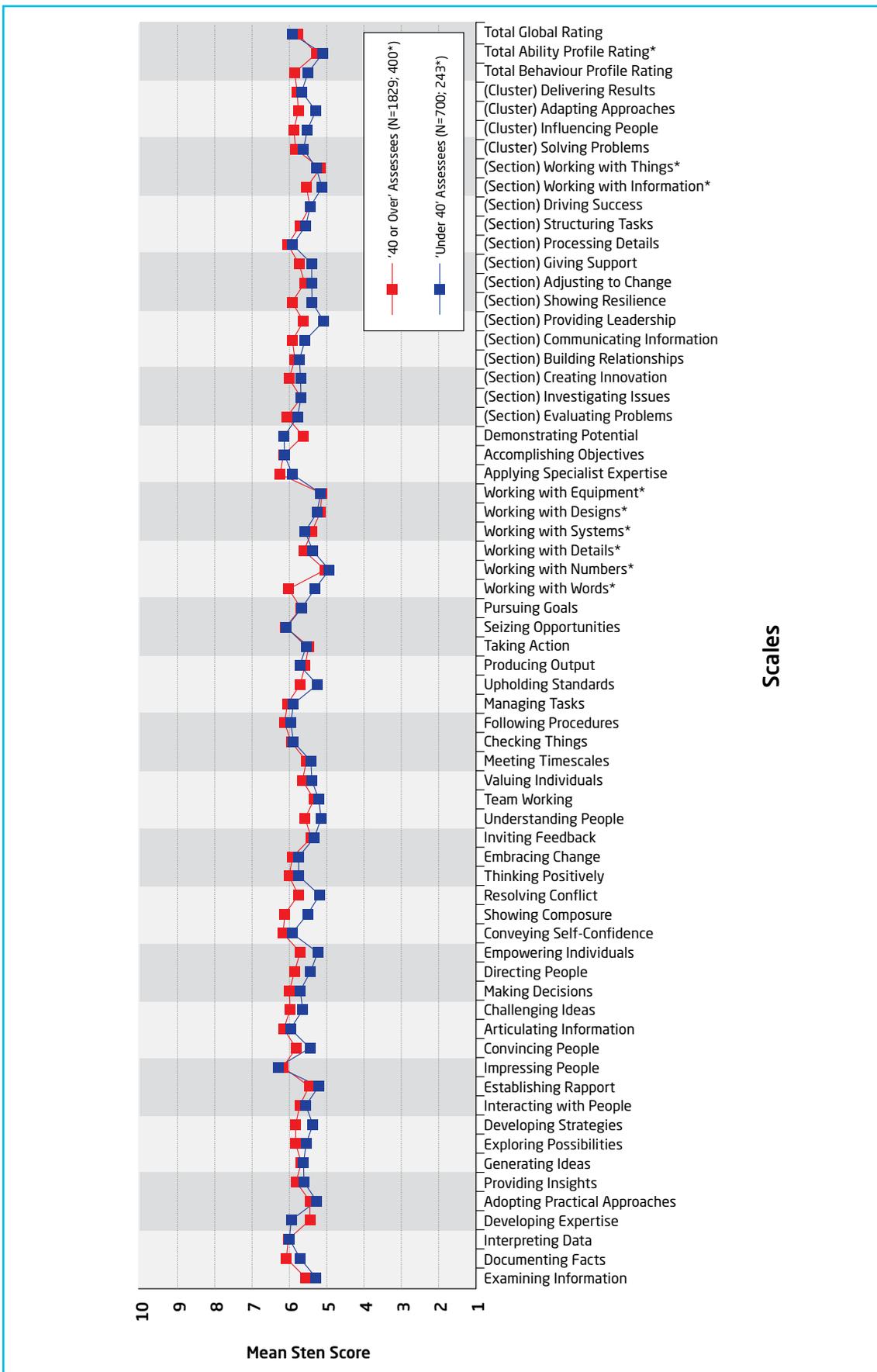


Scales

Figures 16.8 Mean Raw Scores given by Raters aged 40 or Over to Assesseees who are 'Under 40' versus '40 or Over'



Figures 16.9 Mean Sten Scores given by Raters aged 40 or Over to Assesseees who are 'Under 40' versus '40 or Over'



16.3 Gender Trends

Figure 16.10 Mean Raw Scores given by Male versus Female Raters (i.e., not Self)

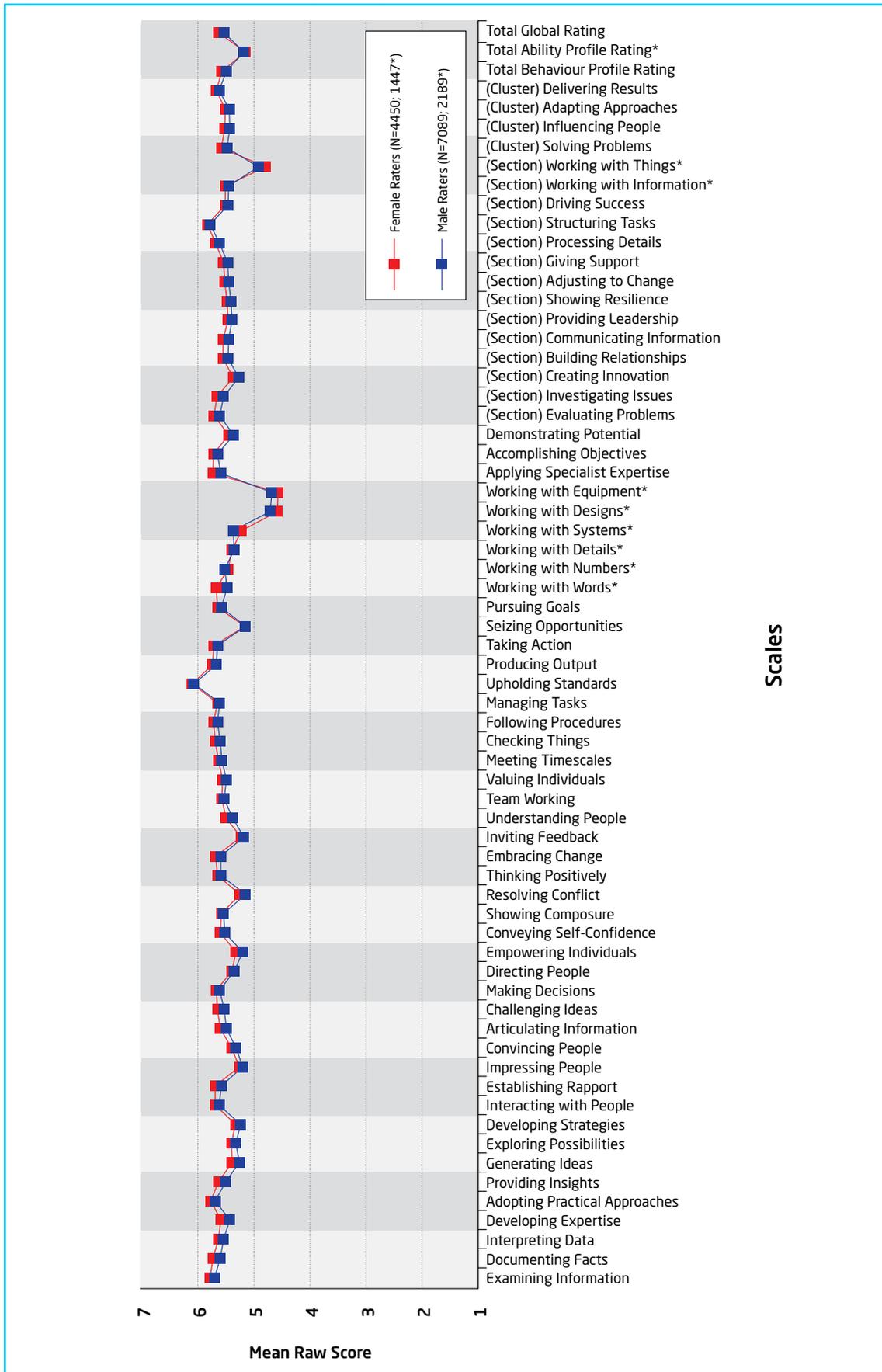


Figure 16.11 Mean Sten Scores given by Male versus Female Raters (i.e., not Self)

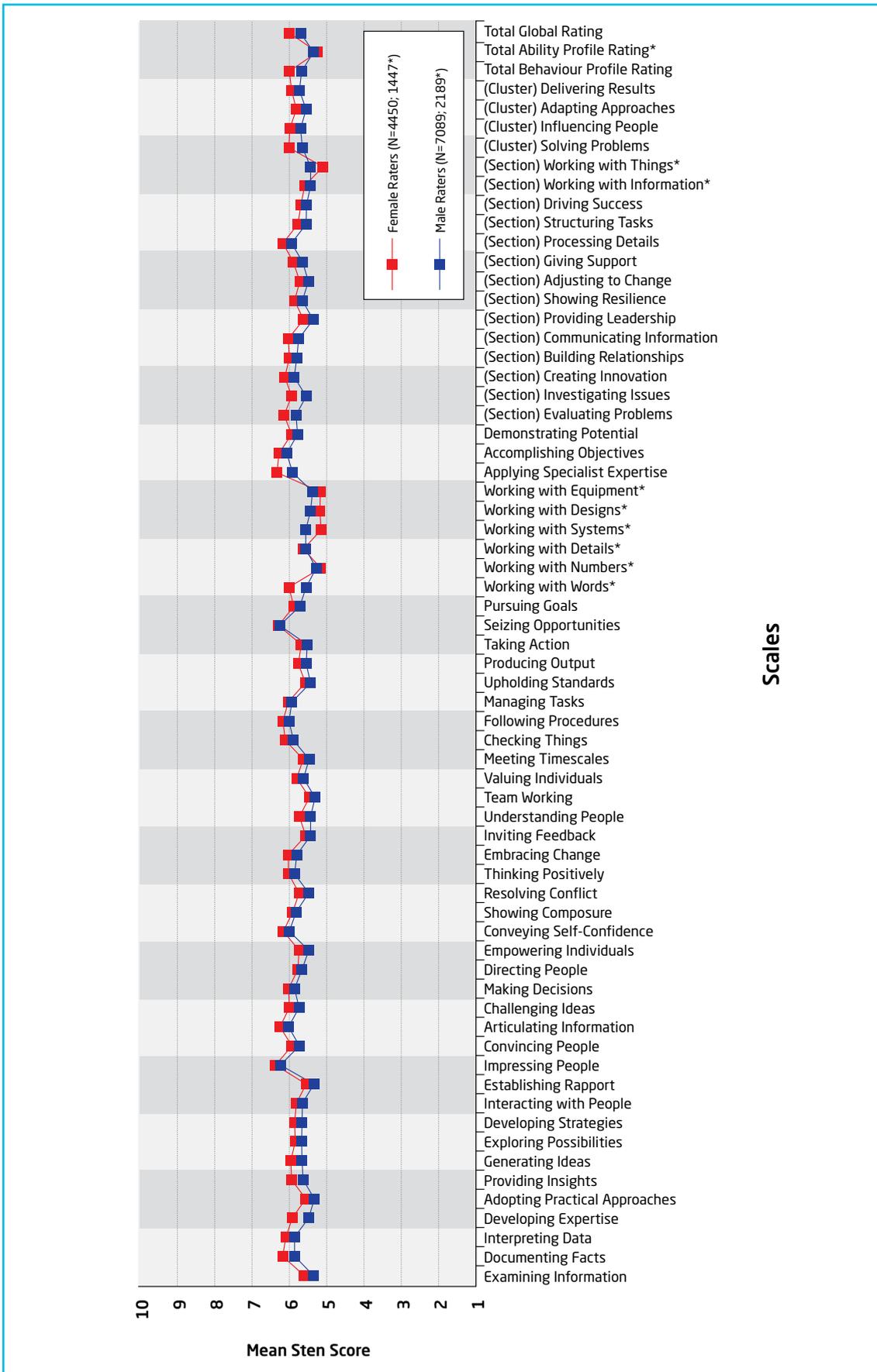


Figure 16.12 Alpha (Internal Consistency Reliabilities) for the Ratings given by Male versus Female Raters

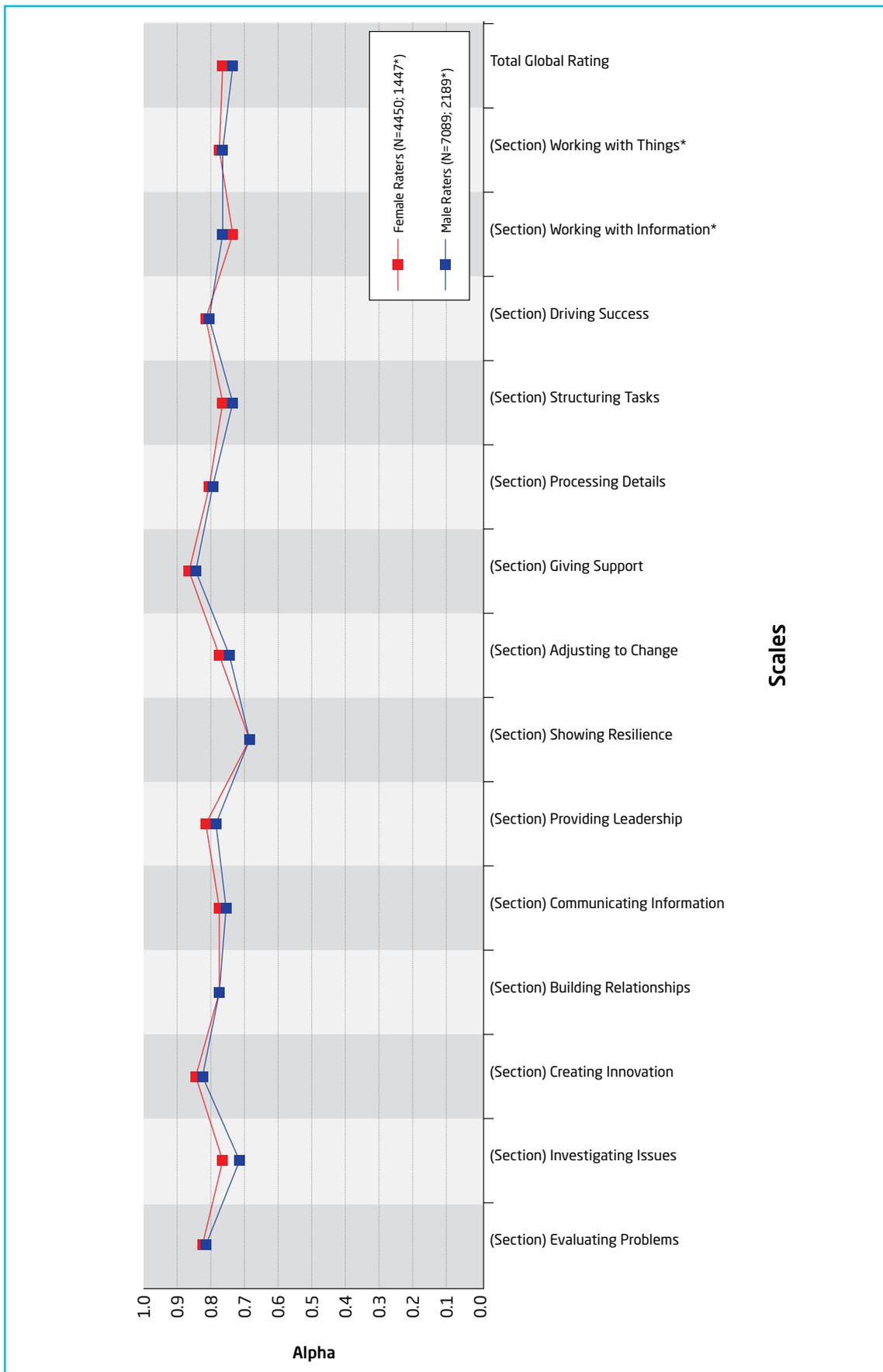


Figure 16.13 Mean Raw Scores given by Raters to Male versus Female Assesseees

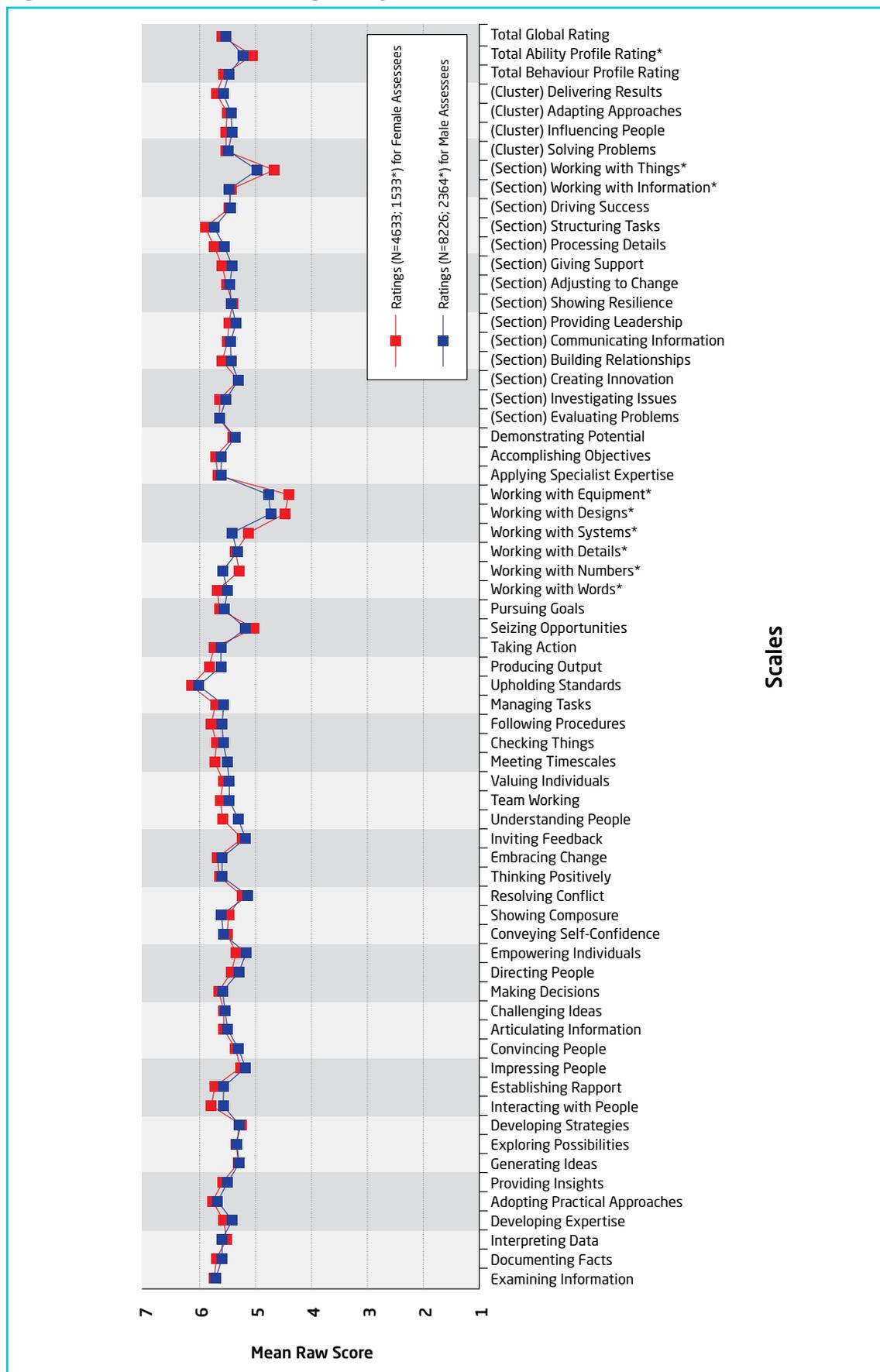


Figure 16.14 Mean Sten Scores given by Raters to Male versus Female Assesseees

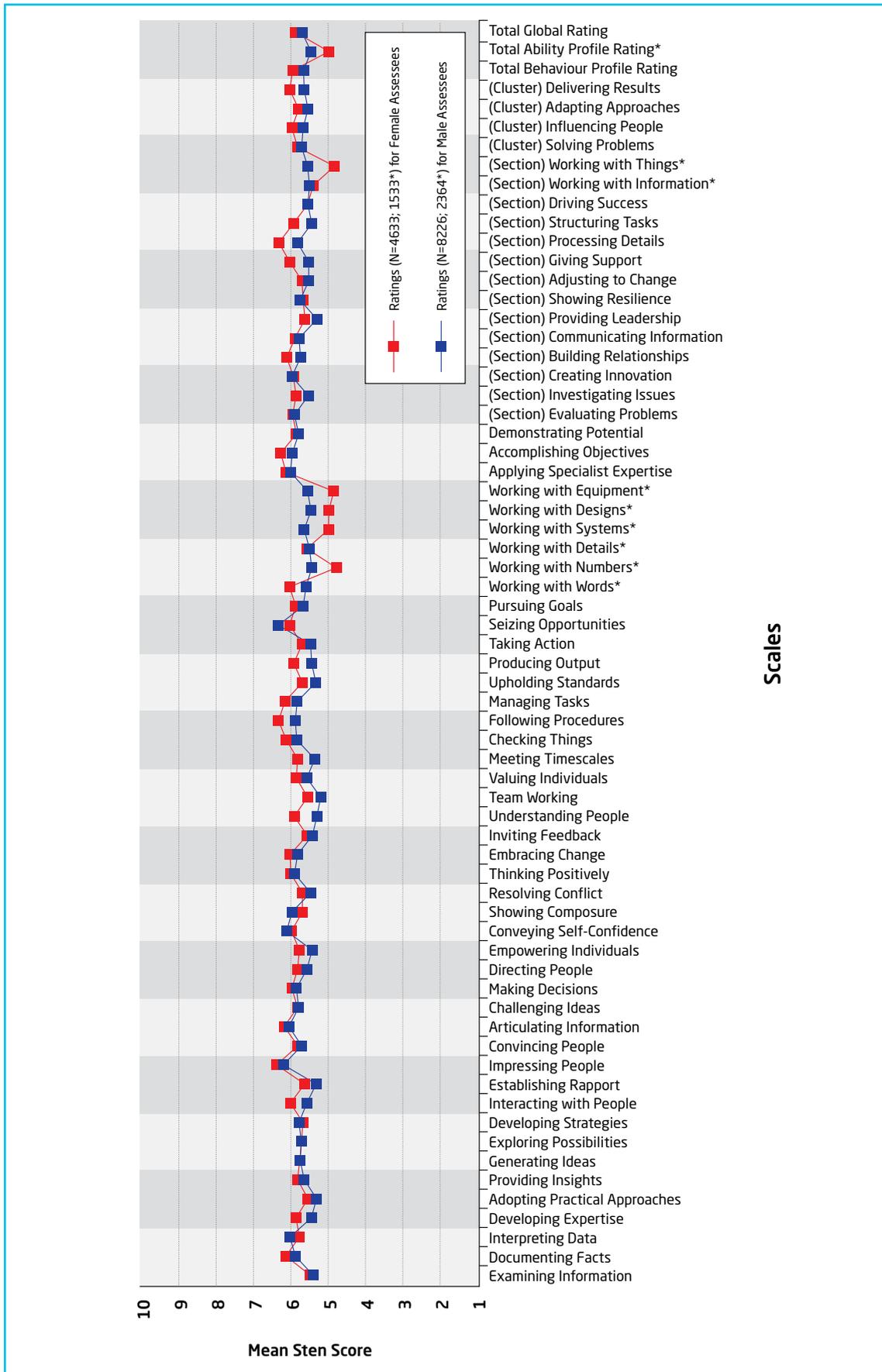


Figure 16.15 Mean Raw Scores given by Male Raters to Male versus Female Assesseees

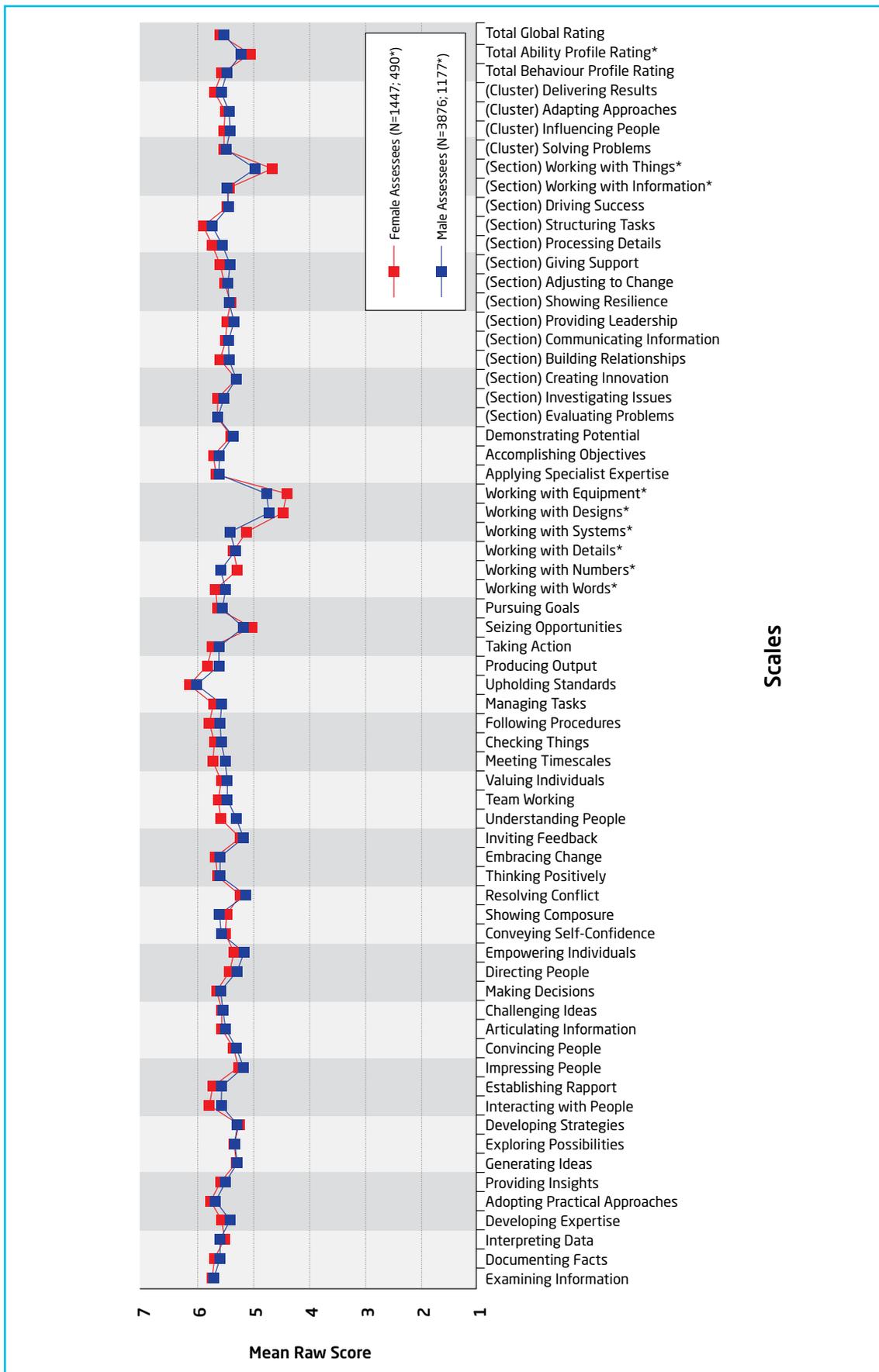


Figure 16.16 Mean Sten Scores given by Male Raters to Male versus Female Assesseees.

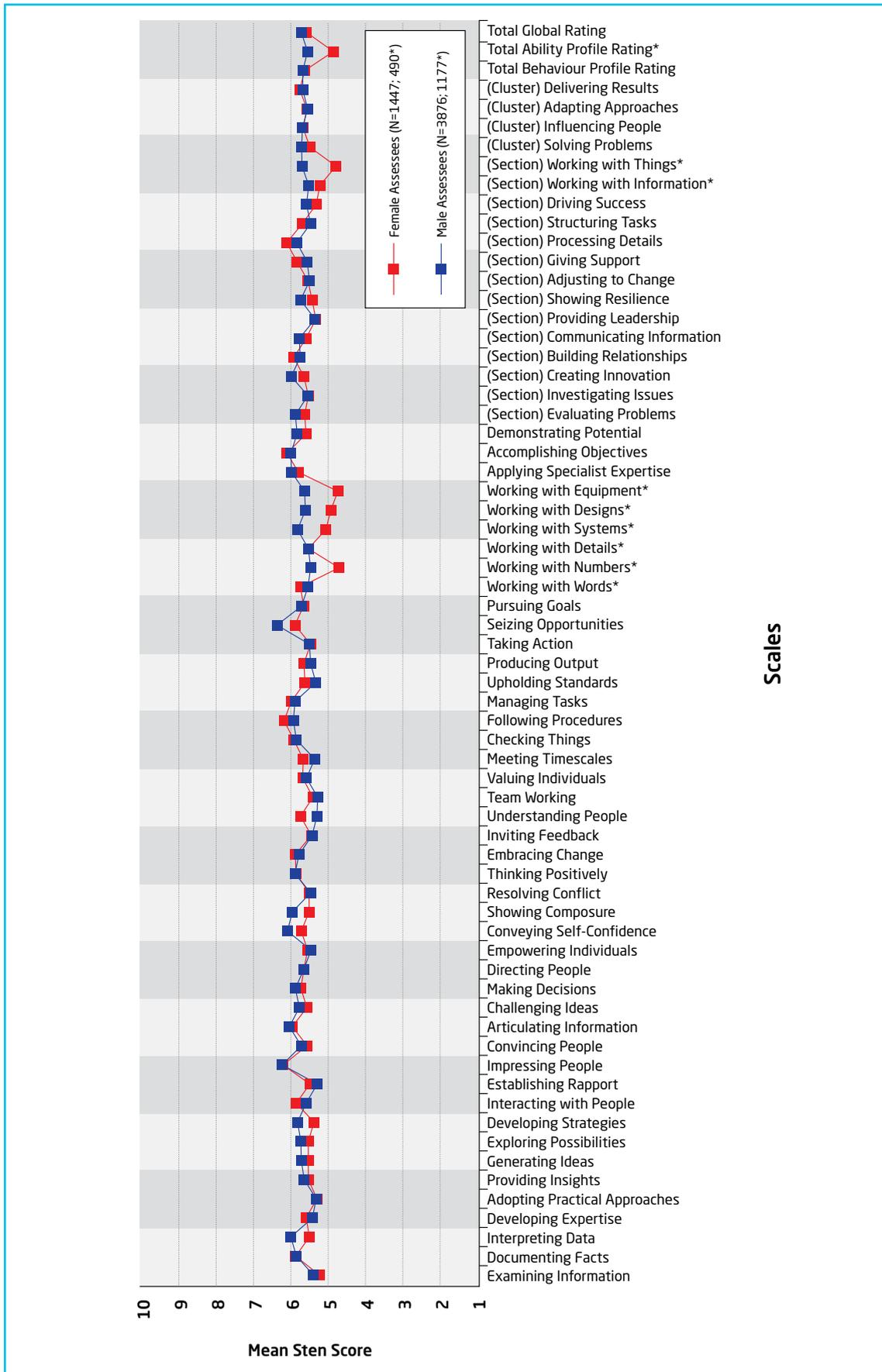


Figure 16.17 Mean Raw Scores given by Female Raters to Male versus Female Assesseees

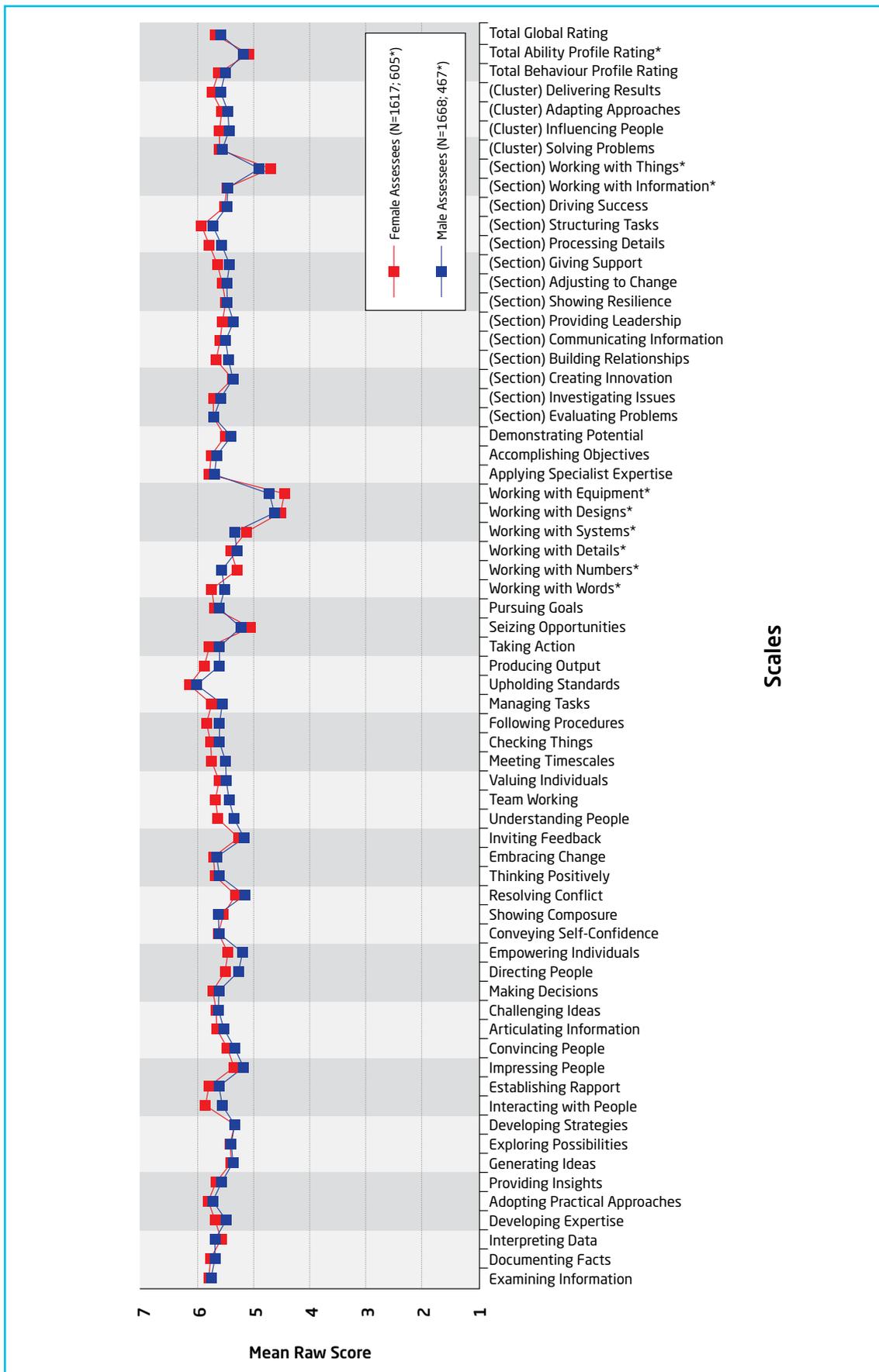
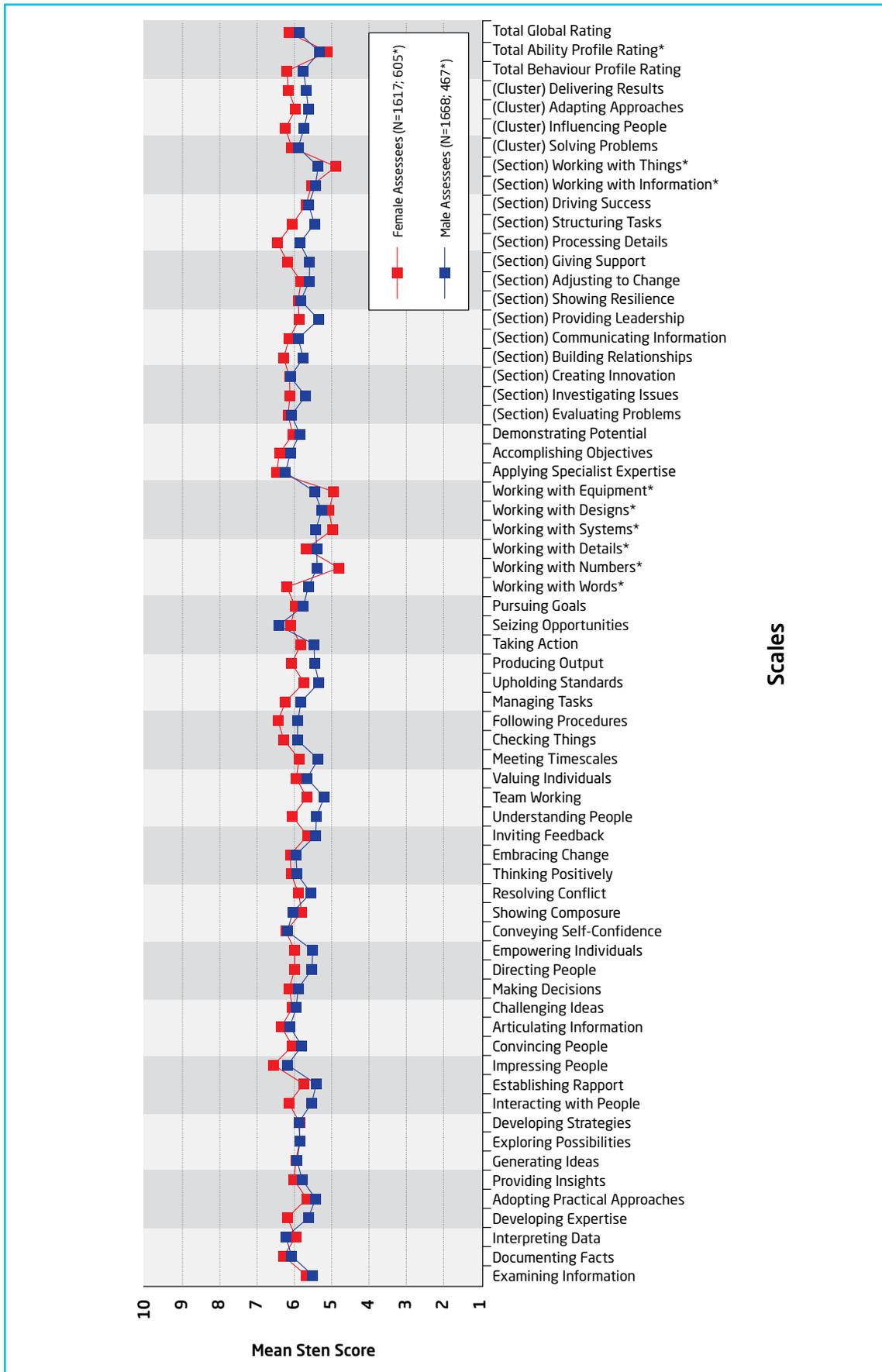


Figure 16.18 Mean Sten Scores given by Female Raters to Male versus Female Assesseees



16.4 Ethnic Background Trends

Figure 16.19 Mean Raw Scores given by UK White raters, Other White raters and those of Other Ethnicities

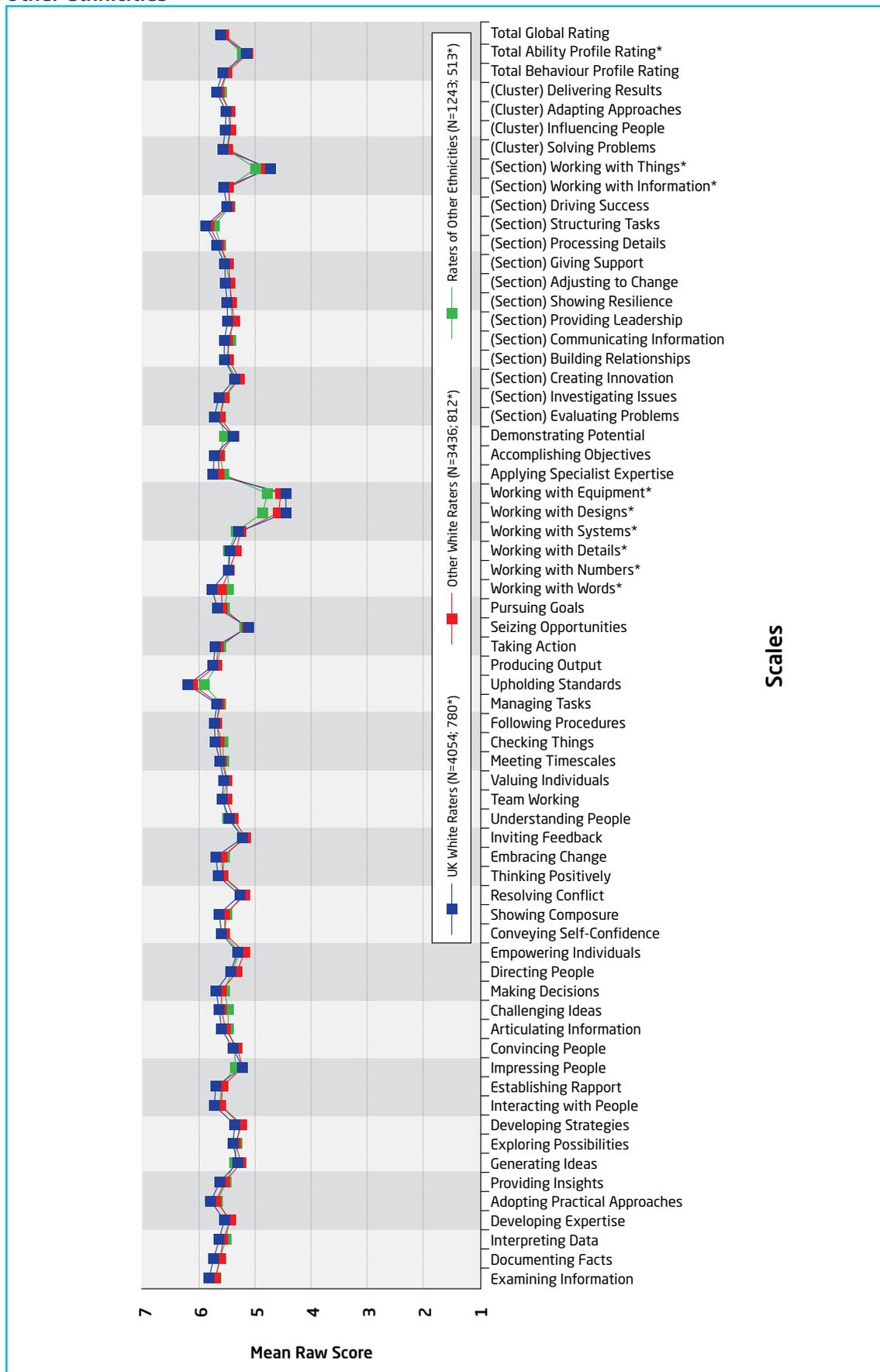


Figure 16.20 Mean Sten Scores given by UK White raters, Other White raters and those of Other Ethnicities

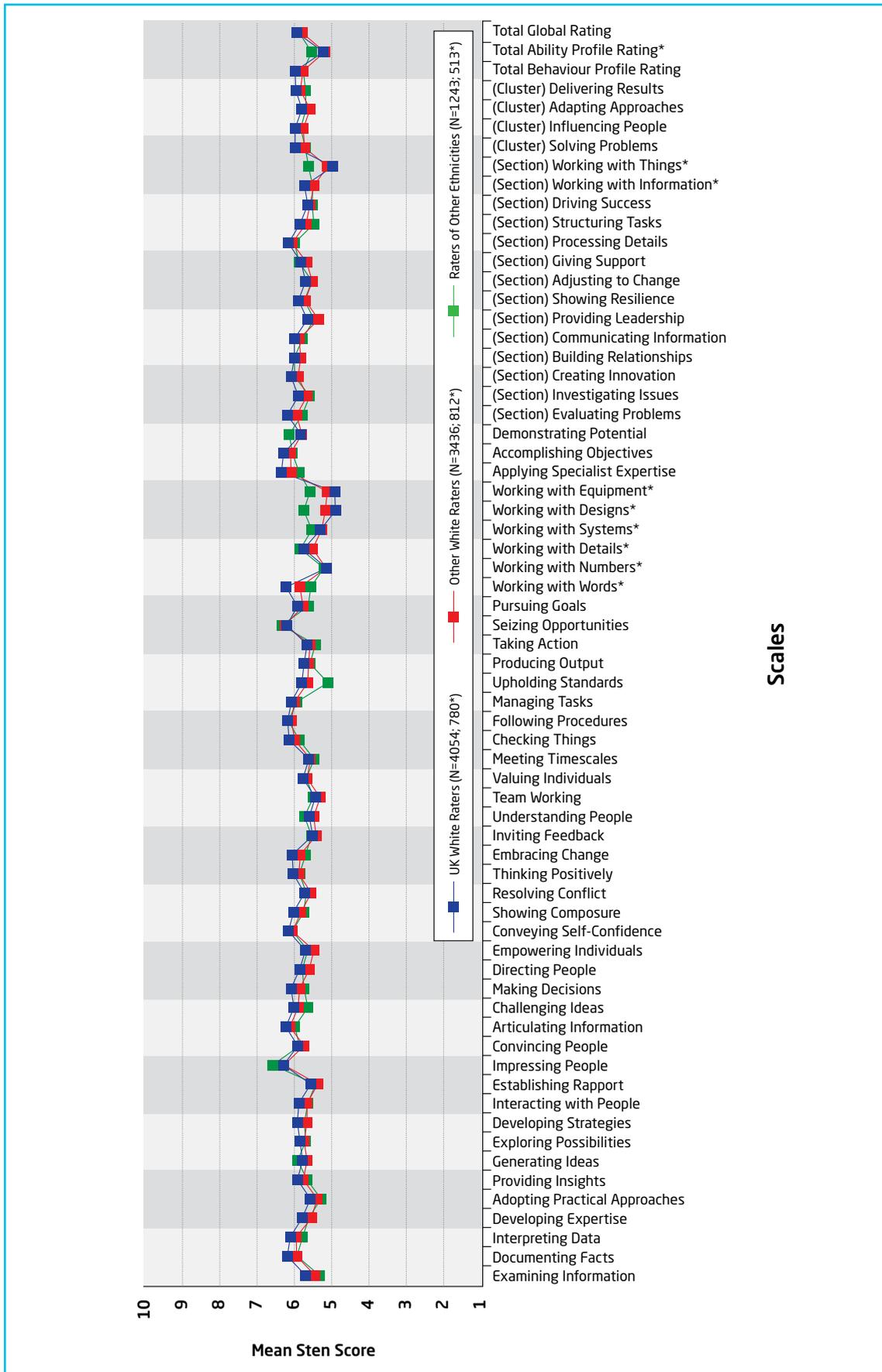


Figure 16.21 Alpha (Internal Consistency Reliabilities) for the Ratings given by UK White raters, Other White raters and those of Other Ethnicities

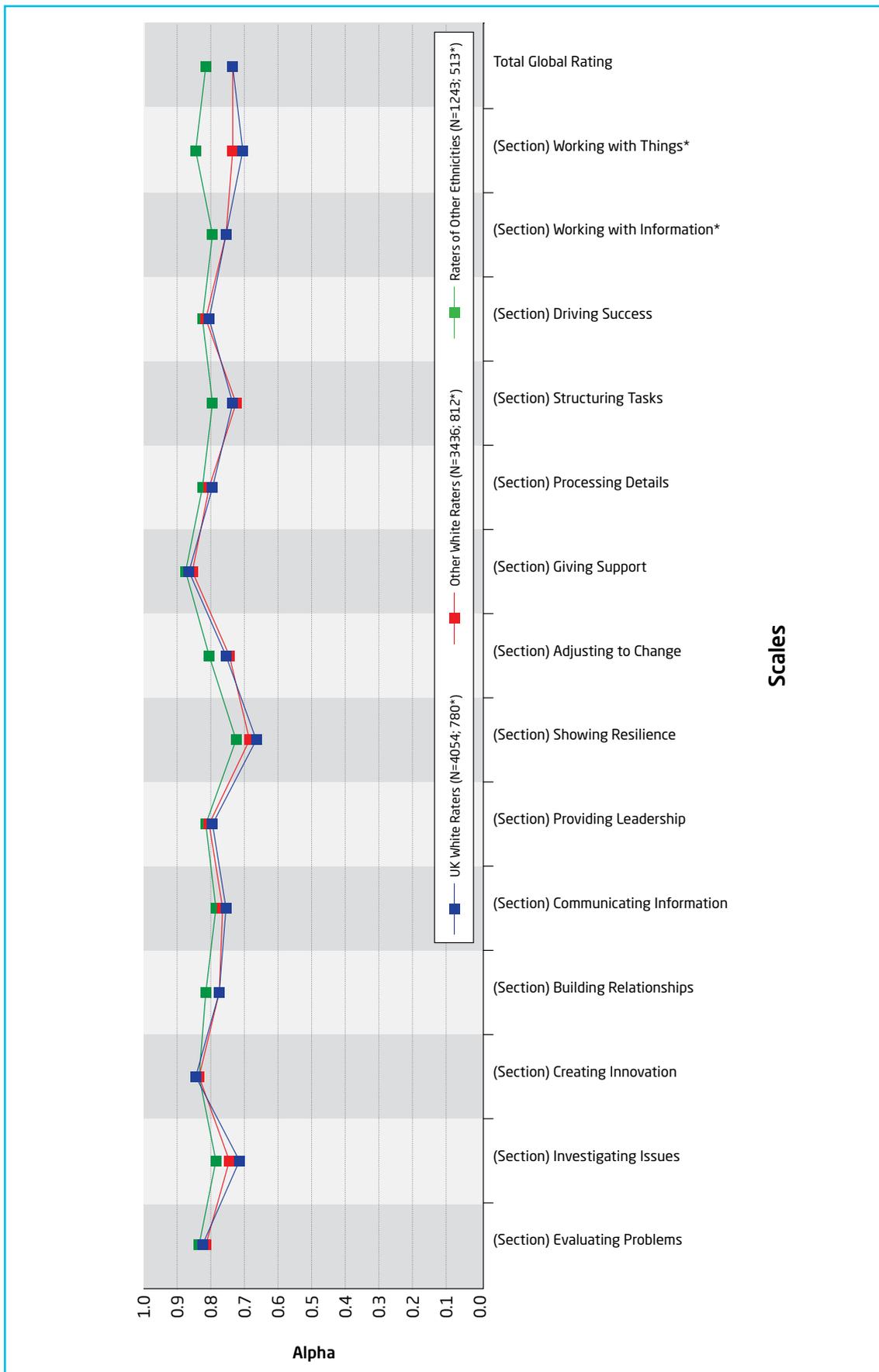


Figure 16.22 Mean Raw Scores given by Raters to UK White and Other White assessees, and those of Other Ethnicities

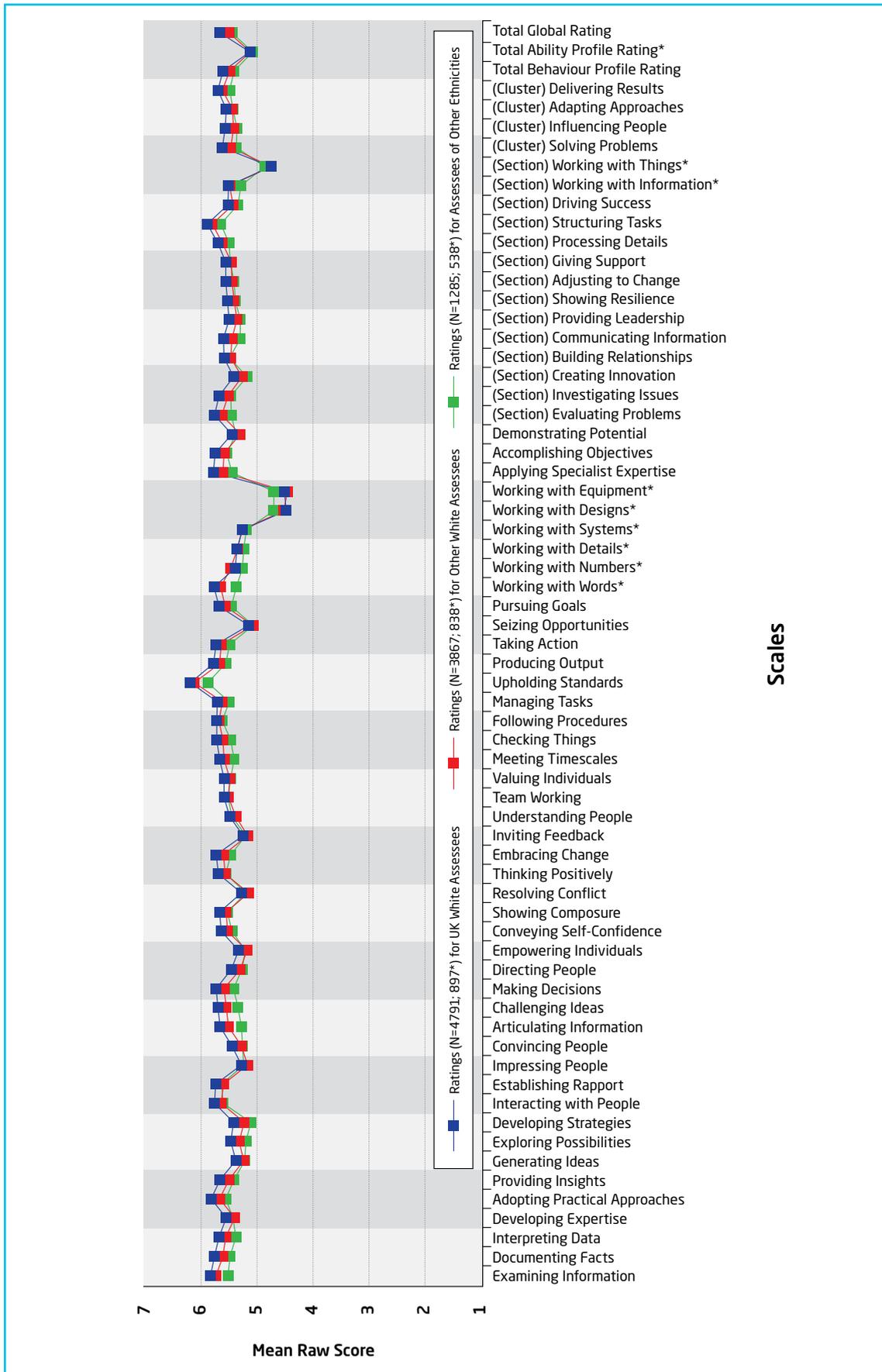
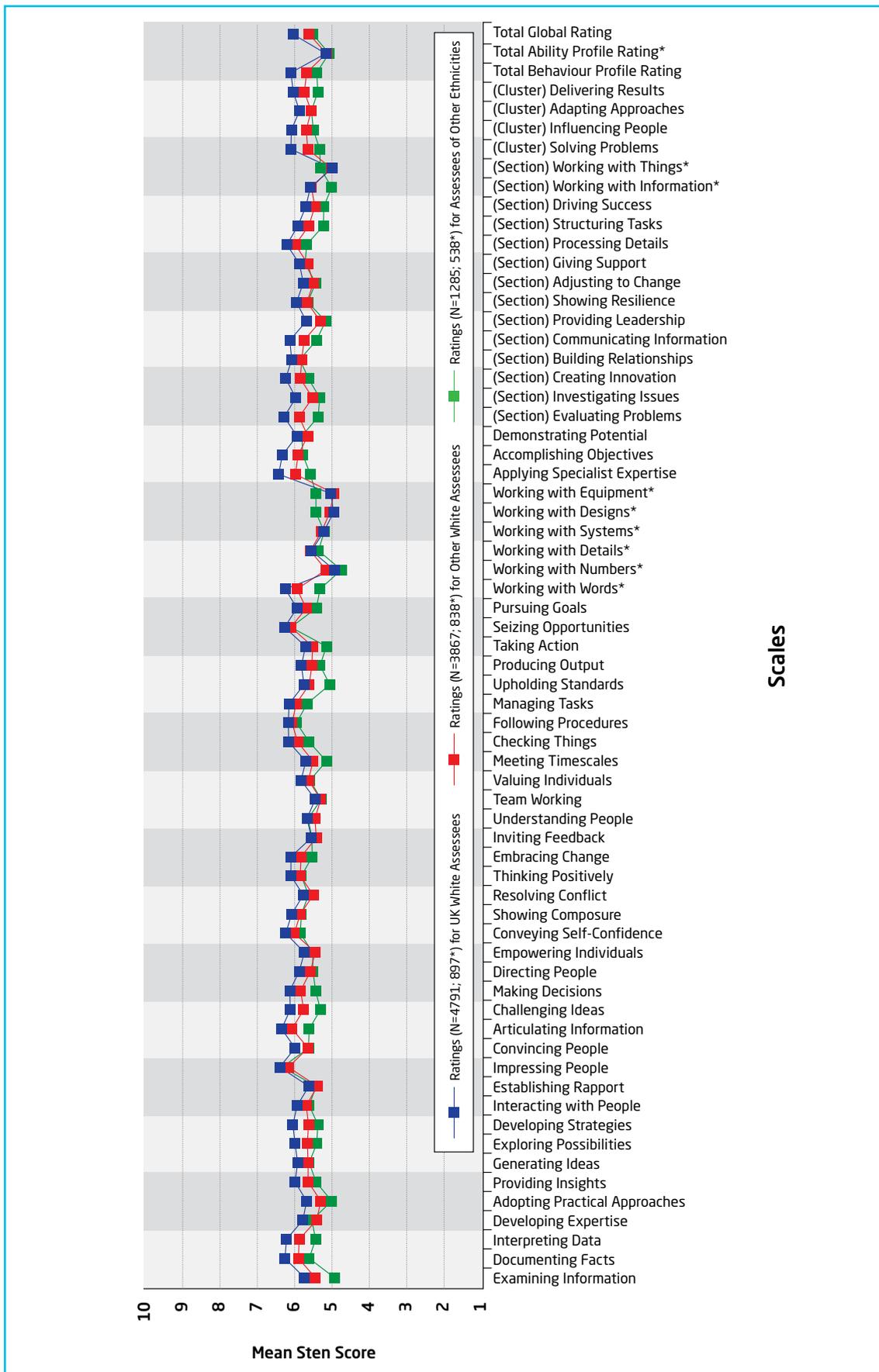


Figure 16.23 Mean Sten Scores given by Raters to UK White and Other White assesses, and those of Other Ethnicities



16.5 Group Trends Descriptions

Age

Figures 16.1 and 16.2 show the mean scores, in raw and sten scores respectively, given by raters under 40 years of age (N=3370; *1192) and those aged 40 or over (N=4268; *1082). When the 'Under 40' group was compared with the '40 or Over' group, none of the 45 scales showed any notable mean difference.

Figure 16.3 shows the internal consistency reliabilities for raters in the two age groups. The mean alpha for the 'Under 40' group was .80 and that for the '40 or Over' group was .78. Differences in the 15 sections were small. This indicates that individuals were reliably assessed using Wave Performance 360.

Figures 16.4 to 16.9 show the mean scores, in raw and sten scores respectively, given by raters for assessees aged less than 40 years and those 40 years or over. A slight trend can be seen across these figures that assessees aged 40 years or above were generally rated slightly higher than those under 40 years.

Specifically, when rated by raters aged under 40 (Figures 16.6 and 16.7), assessees aged 40 or over (N=825; *234) were given higher ratings than those under 40 years of age (N=1205; *503) on Examining Information, Developing Strategies, Challenging Ideas, Making Decisions, Showing Composure, Resolving Conflict, Upholding Standards, Working with Words and Applying Specialist Expertise. The differences in these areas were found to be small (about ½ sten). When rated by raters aged 40 years or above (Figures 16.8 and 16.9), assessees aged 40 or over (N=1829; *400) received higher ratings than those aged under 40 (N=700; *243) on Showing Composure, Resolving Conflict and Working with Words (small differences ranging from .57 to .70 of a sten), whereas assessees under 40 received higher ratings on Demonstrating Potential (small difference of around ½ sten).

Gender

Figures 16.10 and 16.11 show the mean scores, in raw and sten scores respectively, given by male (N=7089; *2189) and female (N=4450; *1447) raters. Female raters seemed to be slightly more generous in their ratings than male raters, although there were no notable mean differences in all 45 scales between the two groups.

Figure 16.12 shows the internal consistency reliabilities for raters in the two gender groups. With the mean alpha for female raters being .79 and that for male raters being .78, differences in the 15 sections were found to be small. This provides further support to the reliability of Wave Performance 360 when assessing individuals.

Figures 16.13 to 16.18 show the mean scores, in raw and sten scores respectively, given by raters for male and female assessees. Small differences can be found between the two groups of assessees.

Specifically, when rated by male raters (Figures 16.15 and 16.16), male assessees (N=3876; *1177) in general were given higher ratings than female assessees (N=1447; *490) on Working with Numbers, Working with Systems, Working with Designs and Working with Equipment (differences ranged from .69 to .88 of a sten). When rated by female

raters (Figures 16.17 and 16.18), female assesseees (N=1617; *605) were given higher ratings than male assesseees (N=1668; *467) on Interacting with People, Understanding People, Meeting Timescales, Following Procedures, Producing Output and Working with Words (differences were about ½ sten), while male assesseees were rated higher than their female counterparts on Working with Numbers and Working with Equipment (differences were about ½ sten).

Ethnicity

Figures 16.19 and 16.20 show the mean scores, in raw and sten scores respectively, for raters in the three different ethnic background groups. The UK White group (N=4054; *780) was compared to Other White (N=3436; *812) and separately to raters of Other Ethnicities (N=1243; *513) in terms of the mean scores they gave. No notable mean difference was found between the ratings given by UK White and Other White raters on all scales. There were small differences found between UK White raters and those of Other Ethnicities in four out of 45 scales, Upholding Standards, Working with Words, Working with Designs and Working with Equipment. UK White raters generally gave higher ratings on Upholding Standards and Working with Words (differences were .69 and .66 of a sten respectively) while raters of Other Ethnicities gave higher ratings on Working with Designs and Working with Equipment (differences were .85 and .66 of a sten respectively).

Figure 16.21 shows the internal consistency reliabilities for raters in the three ethnicity groups. The average internal consistency reliability was highest for raters of Other Ethnicities (mean=.82; highest on 14 of the 15 sections), followed by Other White (mean=.78) then lastly UK White (mean=.77; highest on Creating Innovation). The differences were considered to be small. The results provide reassurance about the reliability of the Wave Performance 360 Sections across different ethnicities.

Figures 16.22 and 16.23 show the mean scores, in raw and sten scores respectively, given by raters for assesseees of the three different ethnic background groups. When the mean ratings received by UK White assesseees (N=4791; *897) were compared to those received by Other White assesseees (N=3867; *838), no notable mean differences were found. When comparing the UK White group with those of Other Ethnicities (N=1285; *538), small differences were found in 23 of 45 scales. UK White assesseees generally were rated higher than those of Other Ethnicities on Examining Information, Documenting Facts, Interpreting Data, Adopting Practical Approaches, Providing Insights, Exploring Possibilities, Developing Strategies, Articulating Information, Challenging Ideas, Making Decisions, Embracing Change, Meeting Timescales, Checking Things, Managing Tasks, Upholding Standards, Producing Output, Taking Action, Pursuing Goals, Working with Words, Applying Specialist Expertise and Accomplishing Objectives (differences ranged from ½ sten to just under 1 sten), whereas assesseees of Other Ethnicities were rated higher on Working with Designs and Working with Equipment (differences were .46 and .39 of a sten respectively).

16.6 Group Trends Summary

The data presented in this chapter on the differences between the means for different groups generally show no or small differences between groups. In each case, whether these variations are attributable to differences in the population means of these groups or are reflective of other variables is not readily discernible. However, some interesting observations can be drawn from the data.

The results in age trends show that assessees aged 40 or over tended to be rated higher, by raters of either age group, on Showing Composure, Resolving Conflict and Working with Words (differences were about ½ sten). Although no firm conclusions can be made, such trends may reflect small effects of maturation, experience and/or seniority.

In gender trends, the ratings given by female raters appeared to be slightly more generous and agreeable. It was found that female assessees were rated higher than male assessees on Understanding People and Resolving Conflicts when all raters were combined (differences were about ½ sten). Female assessees were also rated higher on these two scales, together with Interacting People, Meeting Timescales, Following Procedures, Producing Output and Working with Words when rated by only female raters (differences were about ½ sten). However, as male and female assessees did not differ in the Total Behavior Profile Ratings, these small differences are negligible. Another apparent gender trend was that male assessees were generally rated higher by raters of either gender on Working with Numbers, Working with Systems, Working with Designs and Working with Equipment (differences were just over ½ sten), which resulted in a small difference in the Total Ability Profile Ratings (difference was ½ sten). This trend may partly reflect small genuine gender differences as well as the job roles of male assessees in this sample (e.g., technical roles).

Ethnicity trends showed no notable differences between the ratings given by UK White and Other White raters, or the ratings for UK White or Other White assessees. However, small differences were found in both rater trends and assessee trends between the UK White and the Other Ethnicities groups. The differences in the behavior dimensions are generally small through to moderate (1/2 to 1 sten), and the differences in the ability dimensions are small (under 1/2 sten).

Overall, we see variations within a group (age, gender or ethnicity) but the differences between groups in how they are rated on effectiveness are small or non-existent. These differences do not justify treating age, gender or ethnic background subgroups differently. As a result, we do not recommend using different separate norms for these age, gender or ethnic groups, and do not have such norms available on Oasys. On the contrary, they reinforce the case for using Wave Performance 360 fairly by using one group and a consistent method for a particular job across age, gender and ethnicity.

Two norms are available for Wave Performance 360, Professionals & Managers and Senior Managers & Executives. We advise that the choice of norm group should be appropriate to the level of management of the role being considered.

Appendix A

Correlations between Behavior and Ability Dimension Ratings and Applying Specialist Expertise (ASE) Ratings from Boss, Peer and Report categories (N=14,355 - 1,320 assessments which had at least 1 boss, 2 peers and 2 reports; N=3,849^a - 385 assessments which had at least 1 boss, 2 peers and 2 reports)

Dimension	Boss Dimension Rating against...			Peer Dimension Rating against...			Report Dimension Rating against...		
	Boss ASE Rating	Peer ASE Rating	Report ASE Rating	Boss ASE Rating	Peer ASE Rating	Report ASE Rating	Boss ASE Rating	Peer ASE Rating	Report ASE Rating
Examining Information	.45**	.20**	.20**	.22**	.51**	.26**	.20**	.24**	.59**
Documenting Facts	.37**	.19**	.18**	.20**	.48**	.22**	.19**	.22**	.58**
Interpreting Data	.44**	.22**	.15**	.22**	.53**	.20**	.19**	.23**	.58**
Developing Expertise	.40**	.22**	.17**	.17**	.48**	.17**	.18**	.24**	.61**
Adopting Practical Approaches	.36**	.16**	.13**	.15**	.45**	.21**	.15**	.19**	.61**
Providing Insights	.40**	.19**	.16**	.19**	.50**	.22**	.15**	.21**	.64**
Generating Ideas	.35**	.16**	.12**	.14**	.39**	.16**	.15**	.20**	.59**
Exploring Possibilities	.44**	.22**	.17**	.20**	.53**	.20**	.19**	.24**	.66**
Developing Strategies	.40**	.18**	.18**	.19**	.43**	.16**	.15**	.20**	.58**
Interacting with People	.22**	.10**	.13**	.04	.28**	.14**	.08**	.16**	.52**
Establishing Rapport	.18**	.11**	.15**	.01	.26**	.15**	.04	.13**	.48**
Impressing People	.19**	.13**	.10**	.07*	.24**	.09**	.04	.13**	.41**
Convincing People	.32**	.15**	.18**	.10**	.42**	.20**	.09**	.16**	.58**
Articulating Information	.34**	.17**	.13**	.13**	.40**	.22**	.12**	.19**	.58**
Challenging Ideas	.38**	.16**	.13**	.19**	.49**	.21**	.15**	.20**	.55**
Making Decisions	.35**	.11**	.13**	.13**	.43**	.19**	.14**	.20**	.59**
Directing People	.30**	.09**	.17**	.09**	.36**	.20**	.10**	.17**	.58**
Empowering Individuals	.31**	.15**	.20**	.09**	.39**	.21**	.08**	.18**	.61**
Conveying Self-Confidence	.22**	.10**	.10**	.12**	.33**	.14**	.10**	.15**	.47**
Showing Composure	.20**	.10**	.09**	.11**	.32**	.13**	.08**	.14**	.43**
Resolving Conflict	.26**	.11**	.12**	.05*	.33**	.16**	.05	.10**	.51**
Thinking Positively	.21**	.09**	.08**	.09**	.27**	.12**	.09**	.13**	.52**
Embracing Change	.32**	.12**	.08**	.13**	.37**	.13**	.11**	.16**	.56**
Inviting Feedback	.31**	.13**	.09**	.08**	.37**	.16**	.06*	.13**	.55**
Understanding People	.28**	.14**	.15**	.06*	.34**	.19**	.07**	.16**	.57**
Team Working	.37**	.14**	.15**	.06*	.38**	.21**	.06*	.19**	.59**
Valuing Individuals	.29**	.13**	.10**	.03	.36**	.18**	.09**	.18**	.57**
Meeting Timescales	.30**	.14**	.12**	.15**	.39**	.19**	.14**	.15**	.45**
Checking Things	.40**	.19**	.15**	.23**	.48**	.21**	.19**	.17**	.54**
Following Procedures	.33**	.10**	.08**	.16**	.40**	.17**	.11**	.11**	.49**
Managing Tasks	.35**	.10**	.11**	.20**	.42**	.19**	.13**	.16**	.54**
Upholding Standards	.43**	.21**	.20**	.21**	.52**	.25**	.18**	.26**	.60**
Producing Output	.36**	.14**	.12**	.20**	.48**	.22**	.16**	.18**	.57**
Taking Action	.39**	.16**	.15**	.13**	.43**	.19**	.15**	.20**	.60**
Seizing Opportunities	.27**	.05*	.09**	.03	.26**	.09**	.06*	.05	.46**
Pursuing Goals	.40**	.17**	.17**	.15**	.47**	.20**	.14**	.20**	.63**
Working with Words ^a	.44**	.18**	.18**	.22**	.45**	.17**	.18**	.23**	.52**
Working with Numbers ^a	.25**	.05	.05	.14**	.31**	.12*	.04	.00	.32**
Working with Details ^a	.27**	.08	.05	.12*	.30**	.07	.09	.04	.31**
Working with Systems ^a	.24**	.06	-.01	.10*	.34**	.07	.10	.05	.32**
Working with Designs ^a	.11*	-.04	-.11*	-.06	.17**	.00	-.04	-.05	.20**
Working with Equipment ^a	.09	-.06	-.10	-.12*	.13*	-.04	-.01	.01	.23**
Mean	.32	.13	.12	.12	.39	.16	.11	.16	.52
Median	.32	.14	.13	.13	.39	.18	.11	.17	.57
Min	.09	-.06	-.11	-.12	.13	-.04	-.04	-.05	.20
Max	.45	.22	.20	.23	.53	.26	.20	.26	.66

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

Appendix B

Correlations between Behavior and Ability Dimension Ratings and Accomplishing Objectives (AO) Ratings from Boss, Peer and Report categories ($N=14,355 - 1,320$ assessments which had at least 1 boss, 2 peers and 2 reports; $N=3,849^a - 385$ assessments which had at least 1 boss, 2 peers and 2 reports)

Dimension	Boss Dimension Rating against...			Peer Dimension Rating against...			Report Dimension Rating against...		
	Boss AO Rating	Peer AO Rating	Report AO Rating	Boss AO Rating	Peer AO Rating	Report AO Rating	Boss AO Rating	Peer AO Rating	Report AO Rating
Examining Information	.44**	.23**	.21**	.22**	.51**	.25**	.18**	.21**	.59**
Documenting Facts	.38**	.19**	.18**	.16**	.48**	.18**	.15**	.21**	.57**
Interpreting Data	.41**	.19**	.15**	.17**	.45**	.17**	.12**	.19**	.57**
Developing Expertise	.37**	.21**	.17**	.13**	.48**	.16**	.16**	.23**	.60**
Adopting Practical Approaches	.44**	.17**	.17**	.17**	.53**	.24**	.14**	.20**	.64**
Providing Insights	.47**	.27**	.21**	.22**	.60**	.27**	.15**	.24**	.69**
Generating Ideas	.37**	.23**	.17**	.12**	.48**	.19**	.13**	.23**	.63**
Exploring Possibilities	.42**	.27**	.18**	.16**	.51**	.19**	.14**	.23**	.64**
Developing Strategies	.46**	.27**	.21**	.20**	.56**	.21**	.16**	.24**	.65**
Interacting with People	.36**	.20**	.21**	.14**	.43**	.19**	.13**	.22**	.55**
Establishing Rapport	.27**	.17**	.17**	.07*	.32**	.15**	.07*	.16**	.47**
Impressing People	.32**	.20**	.18**	.17**	.40**	.14**	.13**	.23**	.52**
Convincing People	.42**	.24**	.23**	.16**	.53**	.23**	.11**	.24**	.65**
Articulating Information	.36**	.17**	.17**	.18**	.46**	.23**	.13**	.24**	.60**
Challenging Ideas	.39**	.17**	.16**	.18**	.49**	.22**	.14**	.20**	.59**
Making Decisions	.48**	.24**	.22**	.23**	.57**	.28**	.17**	.27**	.68**
Directing People	.49**	.25**	.28**	.23**	.58**	.30**	.16**	.27**	.69**
Empowering Individuals	.42**	.23**	.26**	.16**	.52**	.27**	.11**	.24**	.65**
Conveying Self-Confidence	.32**	.17**	.16**	.21**	.43**	.20**	.16**	.18**	.57**
Showing Composure	.26**	.11**	.10**	.15**	.35**	.15**	.07*	.14**	.47**
Resolving Conflict	.33**	.16**	.13**	.11**	.39**	.16**	.08**	.13**	.49**
Thinking Positively	.38**	.18**	.16**	.17**	.42**	.17**	.15**	.17**	.58**
Embracing Change	.47**	.19**	.15**	.23**	.52**	.20**	.17**	.22**	.67**
Inviting Feedback	.38**	.15**	.11**	.11**	.44**	.18**	.09**	.16**	.58**
Understanding People	.31**	.18**	.15**	.08**	.38**	.17**	.08**	.17**	.52**
Team Working	.41**	.20**	.20**	.11**	.48**	.21**	.10**	.21**	.61**
Valuing Individuals	.28**	.12**	.11**	.04	.40**	.16**	.08**	.16**	.55**
Meeting Timescales	.46**	.27**	.21**	.27**	.53**	.24**	.19**	.24**	.57**
Checking Things	.40**	.21**	.17**	.24**	.51**	.18**	.16**	.21**	.56**
Following Procedures	.38**	.15**	.10**	.21**	.47**	.15**	.12**	.18**	.53**
Managing Tasks	.48**	.22**	.19**	.30**	.58**	.24**	.18**	.25**	.64**
Upholding Standards	.41**	.17**	.18**	.16**	.50**	.18**	.13**	.22**	.58**
Producing Output	.51**	.27**	.22**	.28**	.61**	.28**	.21**	.26**	.69**
Taking Action	.58**	.30**	.24**	.28**	.64**	.28**	.21**	.29**	.74**
Seizing Opportunities	.40**	.18**	.17**	.12**	.46**	.19**	.12**	.19**	.59**
Pursuing Goals	.62**	.31**	.25**	.28**	.70**	.28**	.18**	.27**	.76**
Working with Words ^a	.28**	.12*	.10*	.01	.33**	.07	.10*	.18**	.44**
Working with Numbers ^a	.22**	.03	.06	.04	.34**	.07	.03	.10*	.34**
Working with Details ^a	.25**	.01	.01	-.01	.27**	-.02	.03	.05	.33**
Working with Systems ^a	.21**	-.02	.00	.02	.29**	.02	.09	.06	.35**
Working with Designs ^a	.13**	-.12*	-.13**	-.02	.16**	.02	-.02	.02	.19**
Working with Equipment ^a	.10	-.10*	-.08	-.12*	.10	-.05	.00	.04	.16**
Mean	.38	.17	.15	.15	.46	.18	.12	.19	.56
Median	.38	.19	.17	.16	.48	.19	.13	.21	.58
Min	.10	-.12	-.13	-.12	.10	-.05	-.02	.02	.16
Max	.62	.31	.28	.30	.70	.30	.21	.29	.76

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

Appendix C

Correlations between Behavior and Ability Dimension Ratings and Demonstrating Potential (DP) Ratings from Boss, Peer and Report categories ($N=14,355 - 1,320$ assessments which had at least 1 boss, 2 peers and 2 reports; $N=3,849^a - 385$ assessments which had at least 1 boss, 2 peers and 2 reports)

Dimension	Boss Dimension Rating against...			Peer Dimension Rating against...			Report Dimension Rating against...		
	Boss DP Rating	Peer DP Rating	Report DP Rating	Boss DP Rating	Peer DP Rating	Report DP Rating	Boss DP Rating	Peer DP Rating	Report DP Rating
Examining Information	.36**	.23**	.19**	.22**	.43**	.26**	.17**	.16**	.52**
Documenting Facts	.32**	.19**	.15**	.16**	.38**	.19**	.14**	.13**	.50**
Interpreting Data	.35**	.17**	.13**	.19**	.38**	.19**	.14**	.14**	.52**
Developing Expertise	.38**	.23**	.16**	.20**	.47**	.20**	.21**	.21**	.56**
Adopting Practical Approaches	.33**	.13**	.14**	.16**	.44**	.25**	.15**	.18**	.56**
Providing Insights	.45**	.23**	.19**	.24**	.54**	.30**	.19**	.21**	.62**
Generating Ideas	.42**	.27**	.19**	.21**	.50**	.22**	.18**	.22**	.55**
Exploring Possibilities	.41**	.27**	.16**	.18**	.47**	.21**	.16**	.20**	.54**
Developing Strategies	.48**	.28**	.22**	.21**	.53**	.22**	.20**	.23**	.58**
Interacting with People	.37**	.21**	.20**	.17**	.41**	.22**	.15**	.19**	.50**
Establishing Rapport	.22**	.14**	.18**	.07**	.26**	.15**	.09**	.11**	.40**
Impressing People	.46**	.27**	.20**	.26**	.55**	.25**	.20**	.27**	.56**
Convincing People	.41**	.24**	.23**	.20**	.49**	.25**	.19**	.20**	.59**
Articulating Information	.39**	.21**	.17**	.24**	.48**	.25**	.19**	.25**	.58**
Challenging Ideas	.42**	.19**	.13**	.21**	.47**	.22**	.16**	.17**	.53**
Making Decisions	.44**	.21**	.23**	.27**	.53**	.29**	.19**	.24**	.60**
Directing People	.44**	.22**	.27**	.28**	.53**	.30**	.21**	.24**	.61**
Empowering Individuals	.36**	.17**	.24**	.17**	.41**	.26**	.16**	.19**	.54**
Conveying Self-Confidence	.46**	.26**	.17**	.28**	.53**	.23**	.19**	.27**	.58**
Showing Composure	.25**	.12**	.10**	.14**	.31**	.13**	.10**	.12**	.40**
Resolving Conflict	.27**	.12**	.13**	.07**	.26**	.15**	.08**	.09**	.45**
Thinking Positively	.31**	.18**	.16**	.15**	.40**	.20**	.14**	.16**	.48**
Embracing Change	.42**	.22**	.18**	.23**	.52**	.24**	.15**	.20**	.58**
Inviting Feedback	.33**	.14**	.11**	.08**	.39**	.17**	.12**	.14**	.49**
Understanding People	.21**	.10**	.12**	.05	.25**	.16**	.07**	.10**	.44**
Team Working	.31**	.14**	.18**	.11**	.36**	.21**	.11**	.16**	.52**
Valuing Individuals	.22**	.06*	.06*	.03	.27**	.14**	.07**	.10**	.43**
Meeting Timescales	.34**	.22**	.18**	.22**	.41**	.22**	.15**	.22**	.47**
Checking Things	.32**	.16**	.15**	.21**	.37**	.18**	.13**	.15**	.47**
Following Procedures	.26**	.09**	.09**	.16**	.30**	.13**	.09**	.10**	.44**
Managing Tasks	.37**	.18**	.17**	.25**	.48**	.25**	.16**	.20**	.54**
Upholding Standards	.26**	.09**	.14**	.12**	.33**	.16**	.09**	.12**	.48**
Producing Output	.42**	.23**	.21**	.29**	.54**	.32**	.20**	.24**	.62**
Taking Action	.51**	.28**	.24**	.31**	.61**	.33**	.24**	.29**	.66**
Seizing Opportunities	.41**	.21**	.16**	.19**	.50**	.20**	.21**	.24**	.54**
Pursuing Goals	.52**	.25**	.25**	.30**	.60**	.30**	.23**	.26**	.66**
Working with Words*	.25**	.12*	.10*	.09	.28**	.08	.10	.14**	.39**
Working with Numbers ^a	.25**	.08	.10*	.14**	.30**	.15**	.06	.09	.27**
Working with Details ^a	.23**	.05	.01	.14**	.30**	.05	.07	.04	.27**
Working with Systems ^a	.18**	.04	.04	.10*	.26**	.08	.09	.11*	.31**
Working with Designs ^a	.13*	-.04	-.08	.07	.18**	.03	.06	.09	.20**
Working with Equipment ^a	.11*	-.02	-.07	-.05	.10*	-.02	.04	.05	.20**
Mean	.34	.17	.15	.17	.41	.20	.14	.17	.49
Median	.35	.19	.16	.18	.41	.21	.15	.18	.52
Min	.11	-.04	-.08	-.05	.10	-.02	.04	.04	.20
Max	.52	.28	.27	.31	.61	.33	.24	.29	.66

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

18.0 References

- Barrick, M.R., & Mount, M.K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology, 44*, 1-26.
- Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology, 90*(6), 1185-1203.
- Borman, W.C. (1997). 360° Ratings: An Analysis of Assumptions and a Research Agenda for Evaluating Their Validity. *Human Resource Management Review, 7*, 299-315.
- Conway, J.M. (1999). Distinguishing Contextual Performance from Task Performance for Managerial Jobs. *Journal of Applied Psychology, 84*, 3-13.
- Conway, J.M. (2000). Managerial Performance Development Constructs and Personality Correlates. *Human Performance, 13*(1), 23-46.
- Costa, P.T.Jr., & McCrae, R.R. (1992). Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. *Psychological Assessment, 4*, 5-13.
- Cronbach, L.J., Rajaratnam, N., & Gleser, B. (1963). Theory of Generalizability: A Liberalization of Reliability Theory. *British Journal of Statistical Psychology, 16*, 137-163.
- Cronbach, L.J., Ikeda, H., & Avner, R.A. (1964). Intraclass Correlation as an Approximation to the Coefficient of Generalizability. *Psychological Reports, 15*, 727-736.
- Digman, J.M. (1997). Higher-Order Factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246-1256.
- Fletcher, C., & Baldry, C. (1999). Multi-Source Feedback Systems: A Research Perspective. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (Vol. 14, pp. 149-193). Chichester: John Wiley & Sons.
- Fox, S., & Bizman, A. (1988). Differential Dimensions Employed in Rating Subordinates, Peers, and Superiors. *Journal of Psychology, 122*, 373-382.
- Garvin, J.D. (1995). *The Relationship of Perceived Instructor Performance Ratings and Personality Trait Characteristics of U.S. Air Force Instructor Pilots*. Unpublished Ed.D. dissertation, Graduate Faculty, Texas Tech University.
- Greguras, G.J., & Robie, C. (1998). A New Look at Within-Source Interrater Reliability of 360-Degree Feedback Ratings. *Journal of Applied Psychology, 83*, 960-968.
- Heneman, R.L., Wexley, K.N., & Moore, M.L. (1987). Performance-Rating Accuracy: A Critical Review. *Journal of Business Research, 15*, 431-448.
- Kurz, R., & Bartram, D. (2002). Competency and Individual Performance: Modelling the World of Work. In I.T. Robertson, M. Callinan & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology*. John Wiley & Sons.

McEvoy, G.M., & Beatty, R.W. (1989). Assessment Centers and Subordinate Appraisals of Managers: A Seven Year Longitudinal Examination of Predictive Validity. *Personnel Psychology*, 42(1), 37-52.

Mount, M.K. (1984). Psychometric Properties of Subordinate Ratings of Managerial Performance. *Personnel Psychology*, 37, 687-702.

Murphy, K.R., & De Shon, R. (2000). Interrater Correlations Do Not Estimate the Reliability of Job Performance Ratings. *Personnel Psychology*, 53, 873-900.

Musek, J. (2007). A General Factor of Personality: Evidence for the Big One in the Five Factor Model. *Journal of Research in Personality*, 41, 1213-1233.

Norman, W.T. (1963). Personality Measurement, Faking, and Detection: An Assessment Method for use in Personnel Selection. *Journal of Applied Psychology*, 47(4), 225-241.

Nyfield, G., Gibbons, P.J., Baron, H., & Robertson, I. (1995). The Cross Cultural Validity of Management Assessment Methods. Paper presented at the 10th Annual SIOP Conference Orlando, USA, May.

Orpen, C. (1973). An Empirical Assessment of the Job Performance of High-Level Executives by Means of a Multitrait-Multimethod Matrix. *Psychologica Africana*, 15, 7-14.

Robertson, I.T., & Kinder, A. (1993). Personality and Job Competencies: The Criterion Related Validity of some Personality Variables. *Journal of Occupational and Organisational Psychology*, 66, 225-244.

Saville, P., MacIver, R., & Kurz, R. (2012). *Saville Consulting Wave® Professional Styles Handbook*, 2nd Edition. Saville Consulting Group, Jersey.

Scullen, S.E., Mount, M.K., & Judge, T.A. (2003). Evidence of the Construct Validity of Developmental Ratings of Managerial Performance. *Journal of Applied Psychology*, 88(1), 50-66.

Spearman, C. (1904). "General Intelligence" Objectively Determined and Measured. *American Journal of Psychology*, 15.

Thornton, III, G.C. (1980). Psychometric Properties of Self-Appraisals of Job Performance. *Personnel Psychology*, 33, 263-271.

Thurstone, L.L. (1936). The Factorial Isolation of Primary Abilities. *Psychometrika*, 1(3), 175-182.

Tubré, T., Arthur Jr., W., & Bennett Jr. (2006). General Models of Job Performance: Theory and Practice. In W. Bennett, Jr., C. E. Lance & D. J. Woehr (Eds.), *Performance Measurement: Current Perspectives and Future Challenges* (pp. 175-203). Mahwah, NJ: Lawrence Erlbaum Associates.

Vernon, P.E. (1950). *The Structure of Human Abilities*. London: Methuen.

Viswesvaran, C., Ones, D.S., & Schmidt, F.L. (1996). Comparative Analysis of the Reliability of Job Performance Ratings. *Journal of Applied Psychology*, 81(5), 557-574.