Saville Consulting Wave Professional Styles Handbook

PART 4: TECHNICAL

Chapter 16: Construction

savilleconsulting

16.0 Construction

This chapter describes the development process used to construct Wave Professional Styles assessments. A general background to Wave Professional Styles can be found in the Introduction chapter. Validation was fundamental to the development of the Wave Professional Styles questionnaires, and users can find further details about the unique approach used to validate the questionnaires in the Validity chapter.

16.1 Development Background

Development Goal

The primary assessment goal of Saville Consulting Wave® Styles is to accurately forecast people's performance at work. This is accomplished by maximizing the criterion-related validity with job performance and other work related effectiveness outcomes. In addition to validity, the Wave Styles questionnaires are designed to be fair in application and applicable internationally.

By achieving these goals, the Wave Styles questionnaires are designed to provide real value to organizations through improved workforce productivity and performance and lead to a strong return on initial investment.

The quest for enhanced validity should be the core mission of the test or questionnaire developer. We also believe it is a priority for assessment and test users. Users have the potential to maximize the benefit they provide in practice by using tools where criterion-related validity is higher.

An assessment tool must be easy to use, acceptable to participants, attractive, and be applicable to today's modern workplace. Saville Consulting Wave questionnaire administration, scoring, and reports are designed to ensure that this validity is easily accessible and understood by the user. We do this by providing the user with clear, well-researched links between Saville Consulting Wave Styles scales and measures of workplace competencies. This enables Wave to give a better, more valid indication of overall effectiveness in terms of proficiency and potential for future success than other conventionally developed personality assessments.
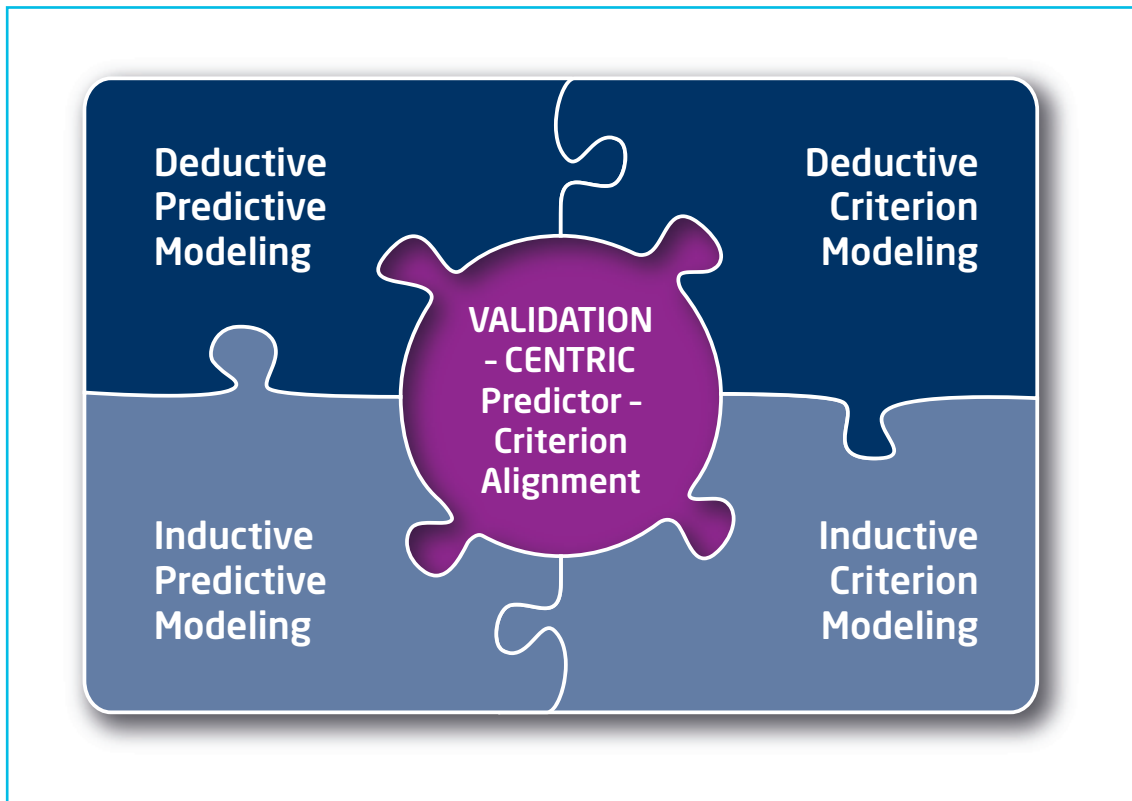
Validation is fundamental to the development of the Wave Professional Styles questionnaires, and users can find further details about the unique approach used to validate the questionnaire in the Validity chapter.

Saville Consulting Integrated Development

Three approaches to questionnaire design were used in the development of Wave Professional Styles. The first two we refer to as "deductive" and "inductive" approaches and they are commonly used to create conventional multi-scale personality questionnaires. The latter approach we refer to as "validation-centric" and it is one of the fundamental design features that makes the Wave Professional Styles questionnaire different.

Conceptual Overview of Wave Integrated Development

Figure 16.1 Conceptual Overview of Wave Integrated Development

Saville Consulting Wave Professional Styles was based on both inductive and deductive modeling at both the predictor and criterion ends of assessment. These predictors and criteria were aligned and the validity maximized by performance driven, validation-centric item/scale choice. Criterion related validity was central to item choice in the development phase of the questionnaire. This is the essence of the validation-centric approach to creating performance driven assessment and maximizing validity.

Inductive Approach to Questionnaire Development

Questionnaires that are inductively developed are based on a process that focuses primarily on its internal structure and not on criterion-related validity. A statistical approach such as factor analysis is commonly used to create personality scales by identifying items that group together (i.e., factors). This approach tends to produce questionnaires that have low correlations between the scales within the questionnaire. Items (and scales) are typically removed that sit between these factors. A concept like Self Assurance, for example, might be removed from a Five Factor Model (FFM) questionnaire measure because it sits between the factors of Emotional Stability and Extraversion – it is related to both factors. The rationale for including an item (or scale), according to this approach, is not based fundamentally on an item's relevance to an external criterion but rather on how it relates to other items in the questionnaire (that is, the

questionnaire's internal structure). The principle objective of the inductive approach is parsimony (as in the FFM) and not improved criterion-related validity. The examination of validity typically is undertaken after the questionnaire is built.

Deductive Approach to Questionnaire Development

Questionnaires that are deductively developed focus on measuring concepts that are deemed relevant to the developer. The developer writes questions to measure a particular concept or trait of interest to them. Developers using a deductive approach do not ignore the internal structure of their questionnaire, however. Typically, the developer writes more items for the trial form of the questionnaire than will be included in the final version of questionnaire (e.g., a third more items than are needed in the final version). The degree to which the items correlate with other items in the same scale is often an important criterion for item selection. Unlike the inductive approach, scales that sit between "factors" will not necessarily be removed from the questionnaire if the developer deems them useful. Questionnaires designed using this approach tend to have scales that correlate more highly with each other when compared to scales developed using an inductive approach. Scales developed using a deductive approach often have high internal consistency reliability, and often upon close inspection, the items in the scale can be found to be repetitive and the concept measured by the scale relatively narrow. As with the inductive approach, an examination of validity typically comes sometime after the questionnaire is built.

Neither the inductive nor deductive method of questionnaire development focuses on empirical validity until late in the development process (or more usually until after the development approach is complete). Both approaches assume that the items written for the questionnaire are good and reliable (i.e., they measure a concept reliably). With both approaches, items are removed that do not correlate with the other items in the questionnaire. But what if those "maverick" items are good predictors of a critical workplace behavior? There is no way of knowing that without external validity data.

Validation-Centric Approach to Questionnaire Development

The validation-centric approach to questionnaire development differs from the inductive and deductive approaches. Item selection begins by looking for items or item groupings (facets) that have the best criterion-related validity rather than selecting items based on how they correlate with other items or how they contribute to a scale's internal structure, such as the item's impact on a scale's mean, standard deviation, or other psychometric characteristics.

An advantage of this approach is the ability to create short and concise but highly valid scales. For instance, Burisch (1997) created short scales by selecting the most valid items from longer scales. Using a double cross-validated research design, he then found that these short scales were capable of greater validity than the conventionally developed scales eight times longer!

Methodological Pitfalls of Validation-Centric Development

A list of methodological pitfalls of a validation-centric approach is presented below along with means of addressing them.

Sample sizes need to be relatively large – recommend in excess of 300 people – to reduce Type I and Type II errors in item selection and to detect differences in validity coefficients. Validation samples of this size are rare as they are typically not easy to achieve early in questionnaire development.

Bizarre items and the black box approach. There is a danger that if item writing and reviewing is not against a clear *a priori*, deductive model that "bizarre" items can be found to correlate with an external criterion. Type I errors can mean that spurious correlations with external criteria can lead to inappropriate items being selected. Items must be written and reviewed against a clear model and principles. A deductive approach to item writing and the use of an *a priori* model can prevent an item intended to measure one concept being used to measure a different concept, e.g., can prevent items about Creativity being used to measure Extraversion. Hence, it is important to make *a priori* hypotheses regarding which items should correlate with which work behaviors and not to rely on a "dust bowl" or "black box" approach to empirical validity.

One validity study does not a questionnaire make. The first draft of a validation-centric developed questionnaire uses the results of an initial development trial to select items with the best validity. However, the results of this initial trial need to be cross-validated at the standardization stage of development as the validation results from the initial sample could be partially inflated by capitalizing on sample specific error variance. Cross-validation ensures the item validities with their aligned criterion are strong and stable across samples of participants.

Performance criteria are inherently unreliable, so it is important that the measures are of clear observable aspects of work performance that can be accurately rated by a boss or co-worker. Single item descriptions of work behaviors or competencies provide the opportunity to break jobs down into their behavioral elements (although these do have limited reliability as job performance criteria).

See the Validity chapter for a discussion on the criterion problem.

While there are pitfalls in using validity data to develop personality questionnaires, they can be overcome yielding benefits for questionnaire users. More work is required at the early stages of questionnaire development, but with large sample sizes, good item writing and review procedures, a large item pool and a clear *a priori* model of work traits and behaviors, questionnaire developers can improve their forecasting accuracy by using validation data early in the development process.

Validity and Scale Development

Burisch (1984) points out that mixed approaches, which include validity data as part of scale development, are surprisingly rare in questionnaire development.

"Actually this is rarely done, particularly the combination of deductive scale writing and external information for item analysis."

In developing the Saville Consulting Wave Styles questionnaires, a mixed approach was used that included deductive, inductive and validation-centric methodologies.
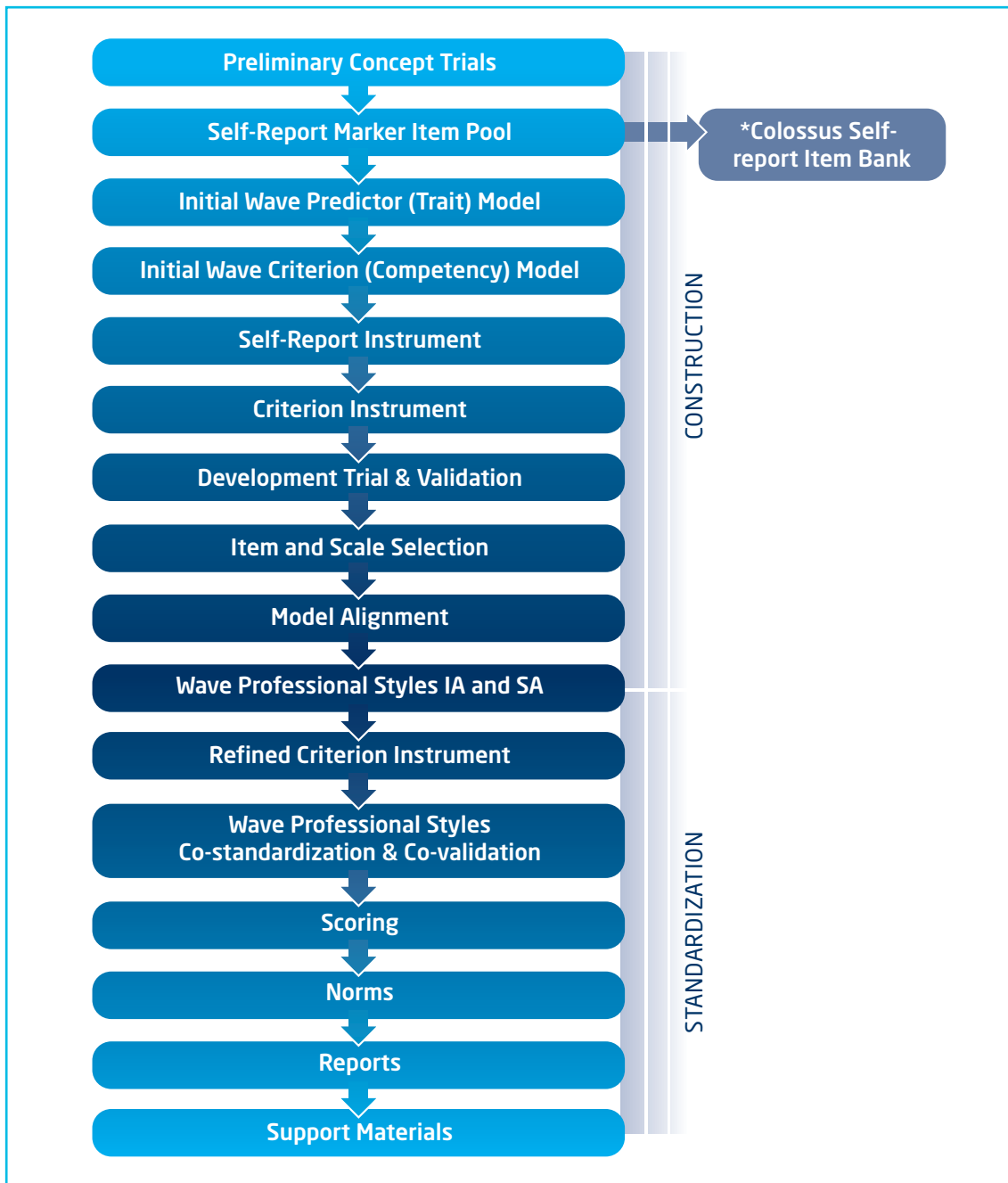
Validity and Expert Judgment

The superiority of validation data in driving expert decision making is a well established finding across different disciplines. The correlation with previous objective criterion/outcome data provides a superior approach to decision making, when compared to the subjective judgment of "experts". This finding is replicated across many different fields of application and is supported by a meta-analytic study by Grove et al (2000).

What is surprising is that the use of criterion-related validity data is so rarely used for item selection and scale development. Most developers develop scales using only internal item parameters. Validation is often an afterthought in questionnaire development - rather than something that drives scale development for the purpose of improving criterion related validity through selecting the most valid items to create more valid scales. Using validity as a rationale for item selection helps to create shorter, more efficient scales with better overall validity.

16.2 Wave Professionals Styles Development

The development of the Wave Professional Styles questionnaires took place in two major phases: the construction phase and the standardization phase. These two phases can be broken down into a number of individual steps as shown in Figure 16.2.

Fig 16.2 Overview of Key Steps in Wave Professional Styles Development

*Colossus Self-Report Item Bank was developed but not used for item/facet scale selection in the Development Trial, the Self-Report Marker Item Pool was used. Parallel items were drawn from Colossus for the Co-Standardization Trial.

Preliminary Concept Trials

In essence, these were "proof of concept and method trials" before investing in the full development and standardization trials of Wave. The content never led to the development of a questionnaire directly from this trial. However, these trials helped to identify the most promising content which was then refined for use in the main development trial.

These preliminary trials were conducted to investigate the properties of different item types; specifically, "effectiveness" items (measuring talent) and "need" items (measuring motive). These trials also pilot tested psychometric properties of a series of short scales (facets) that were designed as part of a model hierarchy to increase the fidelity of assessment by measuring very specific facets of behavior.

A total of over 500 items were trialed on 205 participants from a mixed occupational group. These items were rated against a ten point true/false anchored Likert-type scale. The items were designed to measure 50 different traits. For each of these traits, motive and talent items were written. Items were both positively and negatively keyed. Three alternate forms of the questionnaire were created. Measures of response style and social desirability were included in this trial.

Exemplars of Items included in preliminary trials:

> I **enjoy** generating lots of ideas (motive or need item)
> I **produce** lots of ideas (talent or effectiveness item)

Prior to the study, *a priori* hypotheses were made regarding the structure of the items and scales. These hypotheses were confirmed based on the results of the trials. Specifically, the factor structure and scale intercorrelation matrix reflected the deductive model structure and indicated that it was possible to have a large number of short scales that were internally reliable and that showed good construct independence. A subset of the items was also subject to a retest to check for stability.

The trials confirmed the authors' previous findings that the negatively worded and keyed items, with words such as "not," had less reliability and a less complex factor structure than positively keyed items without negation.

The preliminary trial also provided the necessary information on the scaling properties of the items to begin development of a new dynamic response format. This dynamic response format, designed for use on the Internet, calculates ipsative scores from the rank order of items based on normative ratings, and if certain tied ratings are encountered, will adaptively represent tied items for ranking. The true-false rating scale was amended following this trial to a nine-point Likert-type agreement scale following some small within subject comparative trials of different rating scales.

Self-Report Marker Item Pool

The original Wave development project created 214 self-report concepts or behavioral facets to be used in the development trial (plus measures of impression management and social desirability). These 214 facets were identified through an iterative process of delineation and refinement by a team of six experienced individual difference

psychologists. Each of the 214 concepts was composed of a motive and a talent component making 428 concepts in total (plus social desirability and impression management). In other words, each facet is comprised of two items, one measuring motive and other measuring talent.

An initial draft model was developed by taking the Big Five personality factors and splitting one of the factors, Extraversion, into an interpersonal and achievement-oriented component which is also sometimes considered as part of Conscientiousness. This achievement component differentially predicts important work outcomes which was the reason for separating this out. The resulting six broad general factors were then divided to create twelve categories based on being high or low on each of the six factors.

These twelve categories were further developed to create sub-areas. This was accomplished by examining:

- Criterion frameworks and competency models used by private and public organizations
- Review of existing personality and occupational questionnaires and competency assessment tools
- Validation review of scales from existing assessment tools
- Rational process of splitting scales into meaningful subcomponents
- Identification of new measures, behaviors and performance criteria

Perhaps the most under-rated skill of all in questionnaire development is that of the item writer and reviewer. It is fundamental to Wave Styles' development philosophy that the quality of items is central to the validity of the questionnaire. Over two working person years was taken up with item writing and review by a team of highly experienced individual difference Psychologists.

A set of item writing guidelines and review criteria were created for item writers and reviewers. These are summarized below.

Item Writing Guidelines and Review Criteria

- Targeted
- Simple
- Short
- Unitary
- Unique
- Comprehensible
- Avoiding Idioms and Metaphors
- Non-Opaque
- Self Referent
- Behavioral
- Positively Phrased
- Avoiding Stereotypical, Biased Content
- Non Bizarre
- Non Clinical
- Motive or Talent

462

Targeted – a clear objective in the development of Wave was to write items based on a clear understanding of what "good looks like." Items were designed to be unipolar with higher scores indicating more effective performance. Exemplars of individuals with particular traits were used as part of this process to identify whether each item was worded effectively. For example, one of the facets is labeled "discreet" and its items deal with "maintaining confidentiality." The item writers sought individuals whom they had observed behaviorally to be lower on this trait to check that they would not positively endorse the statement as written. Care was also taken to avoid focusing on concepts that were likely to have negative correlations with overall job performance and potential ratings. Aspects of adapting approaches associated with emotional resilience and agreeableness were, for example, amended to reflect more proactive, pro-social behaviors.

Simple – the items were written and reviewed to be as simple a construction as possible to measure the trait being targeted. Frequently used words were preferred over less frequently used ones. One concept only is measured by an item, and conjunctions such as "and", "or" or "but" are avoided.

Short – long items tend to have lower reliability which will impact on their validity. Wherever possible Wave Styles items were written short and then shortened. However, extremely short or one word items, e.g., an adjective like Sociable, can fail to convey subtlety and precision of meaning when targeting a narrow trait. The longest item in the Wave Styles questionnaires has eleven words and the shortest item has three words. The mean number of words in an item is 6.7.

Comprehensible – an important feature of any item is that reviewers and respondents understand and agree upon its meaning.

Direct – there is no attempt in Wave Styles items to conceal the meaning of any item or make them indirect or opaque. Indeed, they are deliberately written with a high degree of "directness" or transparency so that the respondents are clear what they are being asked and are not surprised later by their results. Opaque or indirect items can lack face validity for the respondent and require many more items in a scale to measure the concept reliably and validly.

Avoiding Idioms and Metaphors – items such as "I like to beat others" can have more than one interpretation. Few idioms or metaphors translate successfully across language and cultures and can lead to the gross misinterpretation of an item. They are best avoided.

Positively Phrased – items that employ negation by using words such as "not" typically have lower reliability and are prone to respondent response error due to misreading. Negative items can help in controlling acquiescence responding, but this can result in lower reliability.

Self-Referent – items ask respondents to rate and rank their own motives and talents. Items that ask for opinions or attitudes such as, "I think a team needs a leader" were avoided.

Behavioral – the items focus on behaviors that are more or less observable. Talent items look for direct expression of these behaviors in terms of action, and motive items focus on the individual's needs and preferences which make particular behaviors more likely. This behavioral approach makes responses more verifiable as the traits should correlate with effective expression of these behaviors which can be observed and rated by others.

463

Avoiding Stereotypical or Biased content – care was taken by reviewers and item writers to avoid items that might focus on specific knowledge or experience which a particular protected group may have less or limited access to. As a matter of course opinions and attitudes on specifics of the real world (such as particular people or events) have been avoided. Items were specifically reviewed by international reviewers with consideration to culture, country of origin, ethnicity, age, gender, sexual orientation, and religious belief. (See Fairness chapter for information about group differences.)

Non-bizarre – item content was not designed to be strange or sensational, or present hypothetical scenarios whether likely or unlikely to be faced by the respondents in their working life.

Non-clinical – items that refer to mental or physical disorders or that describe mental or physical symptoms that may indicate a clinical abnormality were avoided. The items were designed to be relevant and acceptable to typical working adults. Items that could be perceived as an invasion of privacy with regards to mental health issues were also avoided.

Work Relevant - the items in Saville Consulting Wave Styles predictor (trait) model and criterion (competency) model were written and reviewed to be work relevant and positive with regards to achieving overall effectiveness at work.

Internationally Relevant - the items chosen were reviewed to be in simple business English and easy to translate. Individuals from over 50 countries were involved in the item trialing and review process. Words with a common spelling in US and UK English were used whenever possible, but separate versions were created due to a few items containing spelling differences that were not possible to change.

Readability Statistics

Counts		
	Words	1458
	Characters	6550
	Sentences	216
Averages	Words per Sentence	6.7
	Characters per Word	4.3
Averages	Flesch Reading Ease	70.1
	Flesch-Kincaid Grade Level	5.1

Readability Scores

Each readability test bases its rating on the average number of syllables per word and words per sentence. The following sections explain how each test scores your file's readability.

Flesch Reading Ease test

This test rates text on a 100-point scale. The higher the score, the easier it is to understand the document. For most standard files, you want the score to be between 60 and 70.

The formula for the Flesch Reading Ease score is:
206.835 – (1.015 x ASL) – (84.6 x ASW)

where:

ASL = average sentence length (the number of words divided by the number of sentences)
ASW = average number of syllables per word (the number of syllables divided by the number of words)

Flesch-Kincaid Grade Level test

This test rates text on a U.S. school grade level. For example, a score of 8.0 means that an eighth grader can understand the document. For most documents, aim for a score of approximately 7.0 to 8.0.

The formula for the Flesch-Kincaid Grade Level score is:
(.39 x ASL) + (11.8 x ASW) – 15.59

where:

ASL = average sentence length (the number of words divided by the number of sentences)
ASW = average number of syllables per word (the number of syllables divided by the number of words)

Initial Wave Predictor (Trait) Model

A deductive approach to item development was used at this early stage to sort the 214 facets (428 items – a facet is a two-item scale) into the initial model. This was an iterative process that resulted in the framework shown in Figure 12.3.

Figure 16.3 Deductive Facet Styles Bank Structure prior to Item Selection on Development Trial

Solving Problems	Influencing People	Adapting Approaches	Delivering Results
Abstract	Articulate	Accepting	Action Orientation
Analytical	Challenging	Attentive	Commercial
Change	Communicative	Caring	Compliant
Conceptual	Convincing	Calm	Dependable
Creativity	Directing	Customer	Dutiful
Factual	Empowering	Decisions	Energetic
Feedback Criticism	Engaging	Dutiful	Enterprising
Intellectual	Enthusiastic	Ethics	Getting Things Done
Inventive	Gregarious	Facilitating	Insight and Awareness
Learning	Leading	Flexible	Meticulous
Motivating Self	Motivating Others	Insight and Awareness	Organized
Pragmatic	Outspoken	Involving	Planning
Rational	Personal Confidence	Motivating Self	Productive
Thoughtful	Purposeful	Participative	Quality
Visionary	Self-promoting	Positive	Reliable
	Showing Emotion	Receptive	Striving
		Resilience	
		Resolving	
		Self-assured	

Two further scales were developed to measure social desirability and impression management.

Colossus Self-Report Item Bank

Once the initial 428 "marker" items were created, additional parallel items were written to match the original marker items. The result was an item bank of over 4,280 items called Colossus. This item bank was created to allow for the future development of parallel versions of Wave and to be available for custom questionnaire development.

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, the same deductive approach was used to create the Wave criterion model. The goal was to use the model to create items that would be used 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 subcomponents. The 132 competency facet subcomponents in the initial trial were aligned to the deductive Wave predictor model.

Figure 16.4 shows the original criterion model with the clusters and sections. For more information on the underpinning of the structure of the Wave model, see the section later in this chapter on Important Perspectives Related to Wave Model Development.

Figure 16.4 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	Creating Innovation
Implementing Practical Solutions - e.g., applying practical skills utilizing common sense testing ideas	Solving Problems	Creating Innovation
Developing Expertise - e.g., taking up learning opportunities acquiring knowledge and skills exploring specialist issues	Solving Problems	Creating Innovation
Documenting Facts - e.g., finding factsprobing arguments writing fluently	Solving Problems	Creating Innovation
Analyzing Situations - e.g., anticipating problems analyzing information producing solutions	Solving Problems	Creating Innovation
Interpreting Data - e.g., quantifying issues evaluating data objectively utilizing technology	Solving Problems	Creating Innovation
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

468

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	Driving Results
Meeting Timescales - e.g., keeping to schedules; meeting deadlines; finishing tasks	Delivering Results	Driving Results
Following Procedures - e.g., adhering to rules; following instructions; minimizing risks	Delivering Results	Driving 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

Note. 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.

469

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

Note. 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.

Self-Report Instrument

The 428 marker items were placed in an online questionnaire which was used on a Nine-point Likert-type agreement scale (see Administration chapter for further details). The impression management and social desirability items were also included in this trial instrument. The item order was also reviewed to prevent repeats of items from similar constructs presented in a block of items presented at the same time on the computer screen.

Criterion Instrument

A work effectiveness inventory was created using the behavioral items in Figure 12.4 and two items of job proficiency and potential for progression. Ratings were made using 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. The work effectiveness inventories were administered online to a manager, colleague, friend or partner of the Wave Style questionnaire respondent.
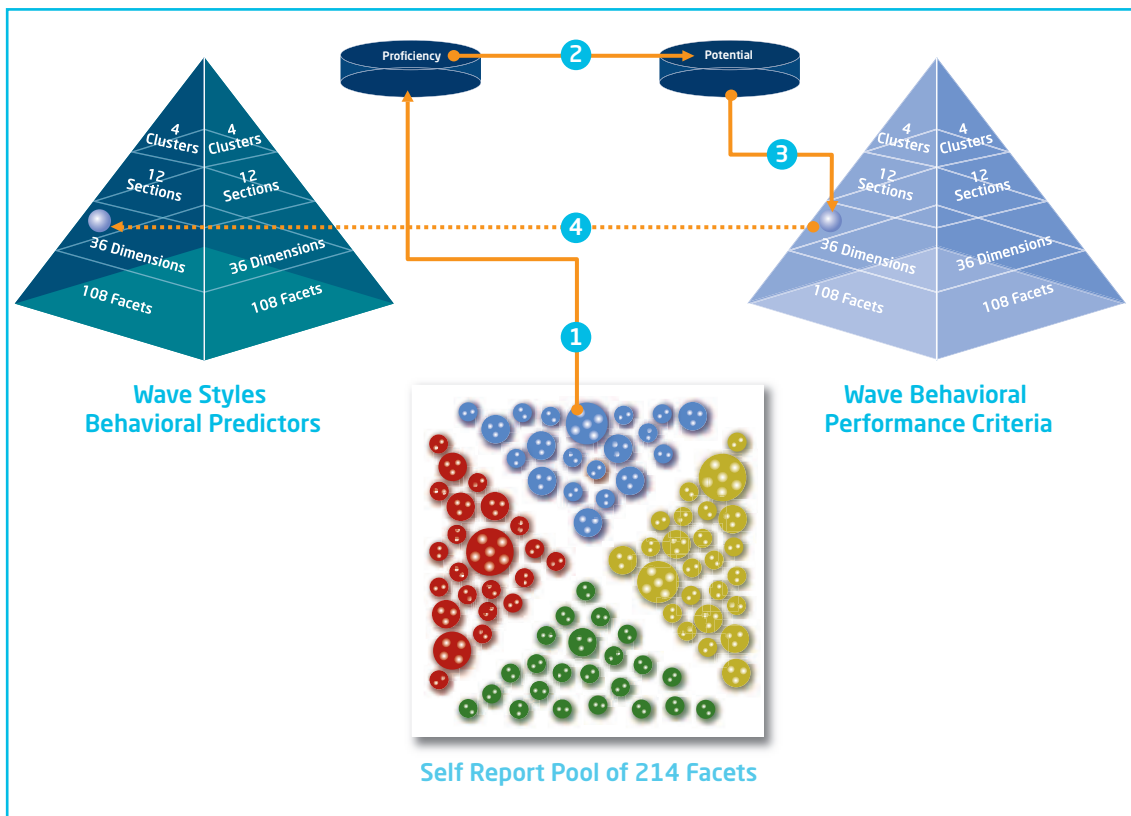
Development Trial & Validation

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 facets (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 16.6 which provides a simplified overview of the item selection process.

Figure 12.6 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 Inventive were examined. The first step was to check that the facets/items correlated with the rater's 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 (potential).

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.

⑥ The new dimension was named based on the content that was selected for inclusion (i.e., the three facets). For instance, the Wave Professional Styles dimension that has been selected to best predict the competency Generating Ideas was named Inventive.

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 weaker 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. This meant that of the original 214 facets only 108 of the most valid facets were selected into the Professional Styles questionnaire. As a reminder, each of the Wave Styles facets is composed of one motive item and one talent item. With 108 facets, this resulted in a 216 item questionnaire.

It should be noted that the ratio of deselected to selected items is much higher than with most self-report questionnaire development. The ratio of deselected to selected Wave items was 1:1 rather than 1:3 which is more typical. This oversupply of high quality items was purposely designed to create the mechanism where validity would be maximized.

There can be a problem with this approach due to the ratio of the number of variables being selected from versus the number of subjects in the study. This problem, can lead to results which do not cross validate as the initial selection capitalizes on error variance or chance outcomes. This was controlled by *a priori* limiting the number of variables that were available to be selected for criteria.

It was interesting to note during item selection that items and facets that correlated strongly with each other in the trials often had far from identical patterns of correlation with the criteria. This is a finding that was also observed in our co-validation research of different instruments, Project Epsom, where scores from assessment instruments that correlate strongly with each other had markedly different correlations with the criteria (see Validity chapter for more information about Project Epsom).

Additional Item Selection Criteria

While the choice of items was based primarily on the validity of the criteria against overall proficiency and potential and of the predictors' validity in relation to the aligned behavioral criterion, other considerations were important in the selection of items into the Wave Professional Styles model.

Again, the first concern was for enhanced validity and not for an optimal factor structure of the predictors and the criteria. Wave Styles was not designed to have a "pure" factor structure. However, there was a need to create predictor measures that could be organized into a hierarchy with a good degree of construct independence while demonstrating acceptable degrees of internal consistency reliability (although alternate form reliability in the subsequent standardization trial provided a more critical test of precision of measurement).

Therefore, once the Wave dimensions had been created using the validation-centric method described, an examination of the internal consistency reliability was undertaken. The goal at the Wave dimension level was to create dimensions with an internal consistency between .60 and .90 (using Cronbach's Alpha). With regards to our goals for test-retest and alternate form reliabilities, our aim was for the Wave dimensions to have the highest possible reliabilities.

Intercorrelation matrices of the Wave facets were run on the 1,011 respondents in the development sample. These were supplemented by a number of factor analyses. These included exploratory factor analysis with a relatively high number of factors. For example, a Principal Components Analysis with Varimax rotation to create 50 factors. This helped to provide an insight of how the facets grouped.

16.3 Model Alignment - The final model of predictor and criterion

Figure 16.7 The Final Matched Model - the top three levels of Wave Professional Styles

Predictor Cluster		Criterion Cluster
Thought	▶	Solving Problems
Influence	▶	Influencing People
Adaptability	▶	Adapting Approaches
Delivery	▶	Delivering Results

Predictor Section		Criterion Section
Evaluative Investigative Imaginative	▶	Evaluating Problems Investigating Issues Creating Innovation
Sociable Impactful Assertive	▶	Building Relationships Communicating Information Providing Leadership
Resilient Flexible Supportive	▶	Showing Resilience Adjusting to Change Giving Support
Conscientious Structured Driven	▶	Processing Details Structuring Tasks Driving Success

Predictor Dimension		Criterion Dimension
Analytical Factual Rational Learning Oriented Practically Minded Insightful Inventive Abstract Strategic	▶	Examining Information Documenting Facts Interpreting Data Developing Expertise Adopting Practical Approaches Providing Insights Generating Ideas Exploring Possibilities Developing Strategies

Predictor Cluster	Criterion Cluster
Interactive	Interacting with People
Engaging	Establishing Rapport
Self-promoting	Impressing People
Convincing	Convincing People
Articulate	Articulating Information
Challenging	Challenging Ideas
Purposeful	Making Decisions
Directing	Directing People
Empowering	Empowering Individuals
Self-assured	Conveying Self-Confidence
Composed	Showing Composure
Resolving	Resolving Conflict
Positive	Thinking Positively
Change Oriented	Embracing Change
Receptive	Inviting Feedback
Attentive	Understanding People
Involving	Team Working
Accepting	Valuing Individuals
Reliable	Meeting Timescales
Meticulous	Checking Things
Conforming	Following Procedures
Organized	Managing Tasks
Principled	Upholding Standards
Activity Oriented	Producing Output
Dynamic	Taking Action
Enterprising	Seizing Opportunities
Striving	Pursuing Goals

Response Style Scale Development

The social desirability and impression management items were dropped following the development trial. There were three primary reasons for the removal of these scales:

1 Social desirability and impression management factored cleanly with a high factor loading on specific personality based factors indicating that they were measuring specific personality variance rather than any broad response style. (For example, Impression Management Positive and Negative scales and the Social Desirability Positive scale factored cleanly with high factor loadings into a factor with other facets about personal organization. Social desirability factored into a factor about openness to change). Even if high scores could be shown to correlate with response distortion on these scales, it would not be possible to separate out which was personality variance and which reflected distortion, i.e., the large degree of construct irrelevant variance severely limited these scales' effective applicability.

2 Overall measures such as acquiescence and consistency from averaging and aggregating differences across similar items gave broader information on response style.

3 Individual information from the difference between ipsative and normative scores at the dimension level in the standardization trial provide more specific information on where social desirable or undesirable responding within a profile is occurring. That is, rather than having a general social desirability scale, it was possible to tap into desirability in responding by comparing ipsative and normative dimension scores and therefore identify specific scales where over and under-rating may be occurring.

Wave Professional Styles – Standardization Trial

The standardization trial was designed to operationalize the Professional Styles model using a new dynamic response format. This involved asking respondents to first rate each item in a block of six items using a free choice, Likert rating scale. Certain tied items are then represented to the respondents and they are asked to use a forced choice response format to select which of the tied items are Most and on some occasions also Least like them. We refer to this dynamic rating-ranking process as the "Ra-Ra" response format. This format allows for both a normative and ipsative score for each Wave dimension. (More information is given on the Response Format chapter.)

In the standardization trial, the 216 items selected from the development trial were used with the new Ra-Ra response format. The items were arranged in blocks of six. In all the 216 items were covered in 36 blocks of six items.

A number of design constraints were used for assigning the items into blocks.

1 Blocks of six were arranged to be composed of entirely motive items or entirely talent items.

2   The structure of the blocks was developed to balance the number of comparisons across the model, e.g., if two dimensions are regularly paired together due to an uneven model then it can create an unrealistically negative correlation between these two scales. It was not possible to accomplish perfect balance in the design using 36 scales, but never the less it was important to ensure a well balanced solution.

3   Four of the items were from different clusters in the model (Wave has four clusters).

4   Where clusters overlap in one block the items come from a different section within the cluster.

5   Assign items with similar mean endorsement values (from the development trial) in each block to control for item attractiveness, which is important when items are being ranked. The first constraints still allowed freedom to balance items with similar endorsement values.

Next, 216 parallel items were drawn from the Colossus Bank that provided for an alternate form version of the Wave Professional Styles questionnaire[1]. This resulted in two alternative form questionnaires being co-standardized with a total of 432 items in 72 blocks. One questionnaire is Professional Styles - Invited Access (IA) and the second questionnaire is Professional Styles - Supervised Access (SA). This allowed for alternate form reliabilities to be calculated.

Respondents were first asked to rate each set of six items online in a block on a nine-point Likert-type rating scale ranging from Very Strongly Disagree (1) to Very Strongly Agree (9). The scoring software immediately calculates the rankings from these rating responses, and presents back to the respondent the items which were tied.

There are some exceptions to this tiebreaking step when the software allows for a small number of ties. This was based on experience that some ties can be tolerated without reducing the reliability of the ipsative scores (similarly other questionnaires such as OPQ32i which is fully ipsative do not require full ranking of scores within a block of items).

The respondent then selects the item which is Most like them and the item which is Least like them. This process continues, if necessary, until the ties are removed. However, if there are no ties then the respondent simply progresses to the next block of six items without the need for a ranking task. (See Response Format chapter for more information.)

Refined Criterion Instrument

An update of the original work effectiveness survey was designed for collecting criterion ratings only on the 36 behavioral work effectiveness areas selected following the standardization trial. The instrument also asked raters to provide ratings of overall effectiveness in terms of job proficiency and potential for future progression. This instrument forms the basis of the Saville Consulting Wave Performance 360 multi-rater assessment.

Wave Professional Styles Co-standardization and Validation

The questionnaire was standardized on a total of 1,153 individuals. Independent external criterion ratings were available on between 556 and 658 of these depending on the criterion dimension (a no evidence option led to the variation in the matched validation sample size). As this was an *a priori* model it allowed for cross-validation into a new sample of respondents. Also, having the two questionnaires provided an opportunity to examine alternate form reliabilities. Details of the results of this study can be found in various chapters of the handbook. Details of the sample can be found in the Appendices.

Scoring

At the final stages of development different scoring regimes were investigated. These included the use of matrix algebra to recreate full sets of rankings from an incomplete matrix of rankings. The ipsative score and normative scores were also looked at in different combinations before arriving at a final process for scoring the questionnaires to ensure both high reliability and validity.

The normative score is based on converting the nine-points on the agreement scale to ranking scores. The rankings are initially determined by examining the rating value assigned to each of the six items. The highest rated item is assigned the highest rank and so forth. Ties are broken by the subsequent force choice ranking task so that each of the six items receives a rank score. Each item therefore receives a normative rating score ranging from one to nine and an ipsative ranking score ranging from one to six. The final overall Wave Styles trait score for any scale at any level in the Wave model is based on both the normative and ipsative scores.

Development of Competency Potential Equations

A final stage was to develop the Competency Potential equations that are designed to maximize the validity of Wave Professional Styles in predicting the competencies in the Wave Competency model.

At each level in the Wave model hierarchy there is one predictor component from Professional Styles questionnaire that is aligned to a specific competency (the Wave Style scale and Wave Competency scale are matched *a priori* and validated empirically). The styles scale (and any subcomponents) were selected on the basis of validity to be the highest individual predictor (or predictors), but secondary predictors (other facets from across the model) do provide incremental validity when predicting competency potential.

The development of the competency potential scales identified these additional predictor elements and gave them prediction weights (lower weights than the matched component which account for the majority of the predicted variance).

These equations were subsequently cross-validated to ensure that the equations are robust and can be generalized to new populations of respondents.

478

Example of Competency Potential Facet Equation

Identifying Business Opportunities (Competency Potential Facet) =	
Business Opportunity Oriented (Styles Facet) x 21	+
Leadership Oriented (Styles Facet) x 4	+
Deciding on Action (Styles Facet) x 3	+
Action Oriented (Styles Facet) x 2	+
Visionary (Styles Facet) x 1	

Preferred Culture

In addition, a series of culture assessments was developed and co-standardized with the Wave Professional Styles questionnaires. This allowed for the modeling and validation of the preferred work culture and environment that individuals would like to work in. This provides the basis of the section in the report on predicted culture/environment fit.

Norms

The initial norms were then created based on the initial standardization sample. Further discussion about norms can be found in the Norms chapter.

Reports

Reports were then built to be straightforward, clear, attractive, and easy to understand. The Expert Bundle was designed to have Line Manager and Personal Reports and provide the expert Wave user with information on the individual's style, motive-talent splits, ipsative-normative splits, facet ranges, predicted culture/environment fit and competency potential.

See Introduction and Reports chapters for further information.

Support Materials

Following the completion of the standardization trial, the internet assessment platform called Saville Consulting Oasys - the Online Assessment System, was finalized and launched. Documentation was made available in the form of training slides and technical documentation on Wave Professional Styles for users.

A Preparation Guide was also developed that could be provided to respondents prior to the administration of Wave. The preparation guide is available as a pdf document free of charge. It can be emailed to candidates in advance of completion of Wave Professional Styles. This can be downloaded from www.savilleconsulting.com.

Since its development for gathering validation evidence, the criterion model and surveys have formed the basis of a variety of new tools that are designed to help Wave Professional Styles be applied more innovatively and effectively. These include:

The Saville Consulting Wave:

- Performance Culture Framework Card Deck
- Job Profiler
- Performance 360

16.4 Important Perspectives Related to Wave Aligned Model Development

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

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 is increasingly referred to in the wake of Positive Psychology 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 questionnaire (Costa & McCrae, 1992) is perhaps the most commonly used FFM questionnaire.

The "Great Eight" Competencies

Kurz & Bartram (2002) expanded the FFM by adding three new categories (Intelligence, Need for Power and Need for Achievement) into the Great Eight (G8) competencies. The G8 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.

480

---

1 As each item in Wave has a specific unique measurement purpose, item level equivalence is something that we are working towards from the design stage.
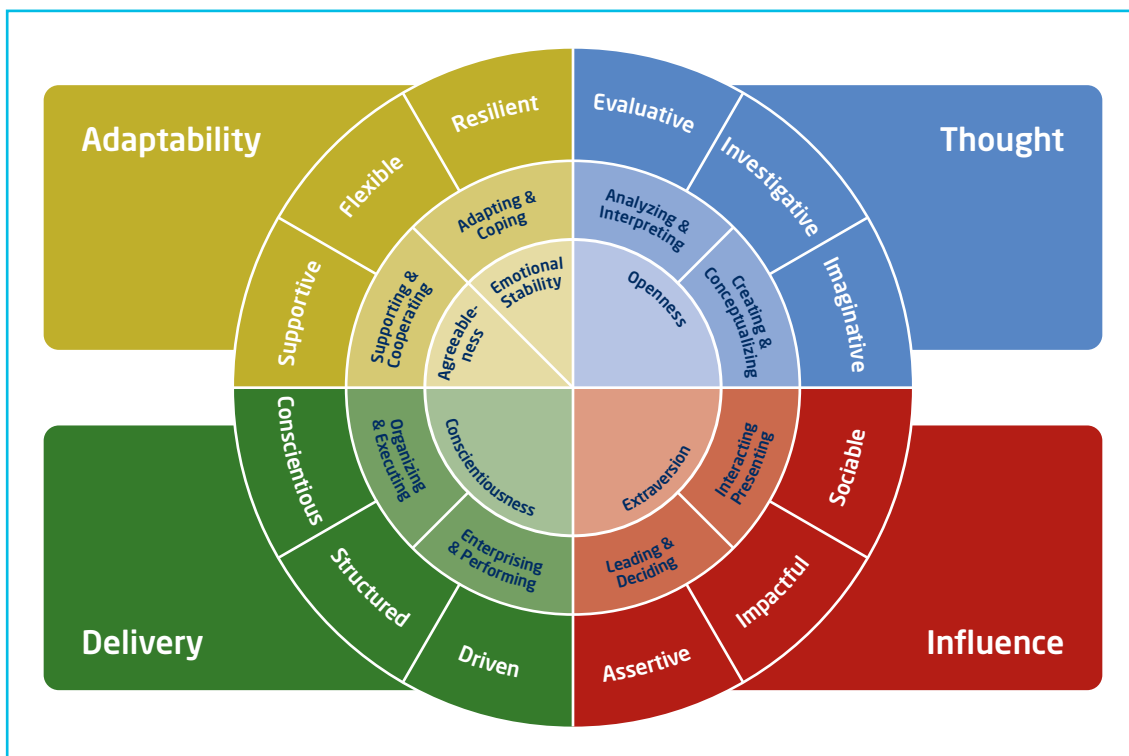
The origins of the model can be traced back to the meta-analysis work of Robertson & Kinder (1993) who created a criterion taxonomy to classify various "local" performance criteria. This led to the development of a range of competency inventories enabling large scale validation work like the international 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.

Alpha and Beta

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 16.8, 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.

Figure 12.8 The Wave Wheel

The Big One

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 super-ordinate 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.

The Four Clusters of Performance and Potential

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.

Bartram, Baron & Kurz (2003) outlined four Leadership Functions that could be viewed as a higher-order structure of the Great Eight. Kurz (2005) measured the Leadership Functions by aggregating relevant competency ratings and identified what ability, interest, motivation and styles variables underpin them. He challenged the distinction of Transformational and Transactional components in the SHL Corporate Leadership Model (Bartram, 2002) as factor analysis results showed that a better account of the data was provided by a top level distinction akin to Digman's Alpha vs. Beta factors in conjunction with the classic People vs. Task orientation well known from leadership research.

The Four clusters of Saville Consulting Wave effectively parallel the four Leadership Functions, but serve as a generic model of work performance across all levels of work complexity and job level.

The clusters are utilized in the Types Report where Influence and Adaptability combine to define four People Types, and Thought and Delivery to define four Task Types. In combination, they define sixteen types that chart transformational and transactional aspects of performance. The Transformer-Transactor type is the most prevalent and effective type in managerial and leadership roles with high staff and resource responsibilities. The Transformer-Transactor aligns with Musek's Big One Factor.

In Saville Consulting Wave there is also matched predictor-criterion alignment and this then forms not only a Big One of personality but also of trait measures that are effective indicators of overall performance or effectiveness at work. The weighting of the factors is based on predictor-criterion associations rather than factor loadings.

Twelve Behavior Sections of Performance and Potential

The Wave model is hierarchical and the four behavioral clusters are each broken down into three sections for a total of twelve 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 complexity of people and jobs.