saville consulting

wave

Saville Consulting Wave Professional Styles Handbook

PART 4: TECHNICAL

Chapter 17: Response Format

saville consulting

17.0 Response Format

The Dynamic Online "Ra-Ra" Response Format

Saville Consulting Wave® Styles use a new dynamic online response format developed and designed specifically for these assessments. The format is neither ipsative nor normative, but a new method which starts with rating and adaptively moves onto ranking when there is insufficient information to differentiate the rated scores into rankings. We call this dynamic rating-ranking process the "Ra-Ra" response format.

The new format is designed to give the benefits of both ipsative and normative formats, while reducing some of the negative consequences of each. We acknowledge that some users may prefer one format over another, and for those that prefer ipsative scoring, they can focus on this score component while those that prefer normative can focus on the normative component.

Before discussing the new format in more detail, a brief overview of some of the properties of ipsative and normative formats are given for reference.

Normative Scores

Normative questionnaires have no inter-relationship between the items; each item stands alone and the response to an item only changes the scores on one primary scale of the questionnaire that the item is a part of.

Simple normative formats include:

Adjective Checklist

Mark all the statements that apply to you by ticking the box next to each statement

- Creative .... ✔
- Reliable .... ☐
- Organized .... ✔
- Adaptable .... ✔
- Talkative .... ☐

Two Answer Option (or dichotomous)

Unipolar - Yes/No or True/False

e.g., I am full of new ideas - True/False?

Bipolar e.g., either/or

Which one of the following best describes you:

Theoretical
Observant

More than Two Answer Options (or polychotomous)

Bipolar

Which of the following best describes you:

I would like to live in the city
In between
I would like to live in the countryside

Unipolar - Agreement Scale

I am creative

| Strongly Disagree | Disagree | Unsure | Agree | Strongly Agree |

Unipolar - Behavioral Frequency Scales

I have ideas

| Never | Sometimes | Frequently | Often | Always |

Unipolar - Behavioral Intensity Scales

My creativity is:

| Low | Moderate | High |

Unipolar- Behavioral Effectiveness Scales

Generating Ideas

| Highly Ineffective | Ineffective | Fairly Ineffective | Unsure | Fairly Effective | Effective | Highly Effective |

N.B. While some of these items above are forced-choice they all can be scored normatively as opposed to ipsatively. The answer options chosen in an item will only impact on the scores on one primary scale of the inventory.

Normative scores have the advantage that individuals are free to be where they want on each and every scale. They could get a maximum score on all the scales (i.e., get a sten score of 10 on each scale) or a minimum score on every scale.

They do, however, suffer from some potential disadvantages:

1. Halo and Horns

As well as raters operating a halo when they rate other people - individuals can view themselves more positively or negatively and self-rate accordingly. This can falsely overinflate or under-estimate an individual's scores on scales of a self-report assessment. This can also falsely lift the intercorrelations in a normative questionnaire where the average intercorrelation across the scales does not reflect the intercorrelations of the true scores. Note. this should be distinguished from an overall true positive intercorrelation that may exist between behaviors (normative scores are likely to provide an overestimate of the true positive relationship between scales on a self report assessment).

2. Central Tendency/Extremity

Individual respondents often differ in the degree to which they use the extremes of polychotomous normative items. This variation in how extreme the responses and hence the profiles are, may be manifest in behavior and be construct relevant. However, it may not be reflected in behavior and be construct irrelevant and thereby misleading to the interpreter of the profile.

3. Acquiescence

Not everyone agrees to statements to an equal extent, and some people are more likely to acquiesce than others. This response pattern could be manifest in an individual's behavior, e.g., being more 'accommodating.' But the response pattern is also likely to impact across the whole profile and it is unlikely a person is 'high' or 'low' across all constructs measured. One approach to reduce the impact of acquiescence is to reverse the polarity of a proportion of the items. However, this may lead to a number of negative consequences such as erroneously bipolar scales, fewer scales created, and use of items with negation (such as 'not'). All of these can serve to reduce fidelity and/or reliability of the assessment.

4 Social Desirability

Social desirability responding can be related to a number of features of individuals, including high self esteem, an eagerness to please, a concern to conform with social norms, etc. The impact of socially desirable responding on a profile can potentially be quite complex. It could have the effect, for example, of the respondent not being prepared to disclose weaknesses which are not considered to be socially desirable, such as 'being inattentive.' Alternatively, it could manifest in a more profound lifting of the scores on 'positive' characteristics across the profile or parts of the profile. Also, the opposite impact on the profile can occur with low social desirability responding.

5. Cognitive Complexity in Self Concept

An associated issue, related to halo and social desirability, is an individual's capacity to differentiate many characteristics within themselves. This capacity can be viewed as an individual difference in itself. Some people may articulate their self concept quite simply and not differentiate the type of extrovert they are in great detail, whereas another individual may have a very detailed, well differentiated, hi-fidelity concept of their extrovert behavior that they can articulate. A normative questionnaire may result in a flat profile for individuals who cannot distinguish characteristics within themselves with great complexity.

6. Faking

Respondents can make a conscious effort to distort the results in, for example, a high stakes selection situation (normally favorably). They also may semi-consciously be trying to create a favorable impression. This can be seen by individuals making the best of what they have rather than being highly critical (analogous in an application form to making the career highlights sound best and minimizing or omitting the negative outcomes, without committing the act of giving inaccurate facts). Such impression management, conscious faking and social desirability are often difficult to differentiate in practice.

Despite all these problems normative questionnaires do work and are valid. However, in high stakes situations there are greater concerns about faking and response distortion. Faking, response distortion and social desirability can reduce opportunities to see the more self-critical responses in many normative profiles.

Ipsative Scoring

With ipsative scoring, the scores have a degree of dependency on each other. Different responses to an item will impact scores on more than one primary scale of the questionnaire. In essence, by giving the scores on an item to one primary scale you take it away from another (or others).

Fully ipsative questionnaires have a fixed total score. If all the scale scores are added together it will always result, by definition, in one fixed value. Therefore, what ipsative scoring is doing is apportioning the scores across the scales.

Ipsative item formats include:

Dichotomous Either/Or

Which one of the following best represents you:

Influential
Organized

Polychotomous Rank

Rank the following statement 1,2,3 and 4, where 1 is most like you and 4 is least like you

> Persuades others
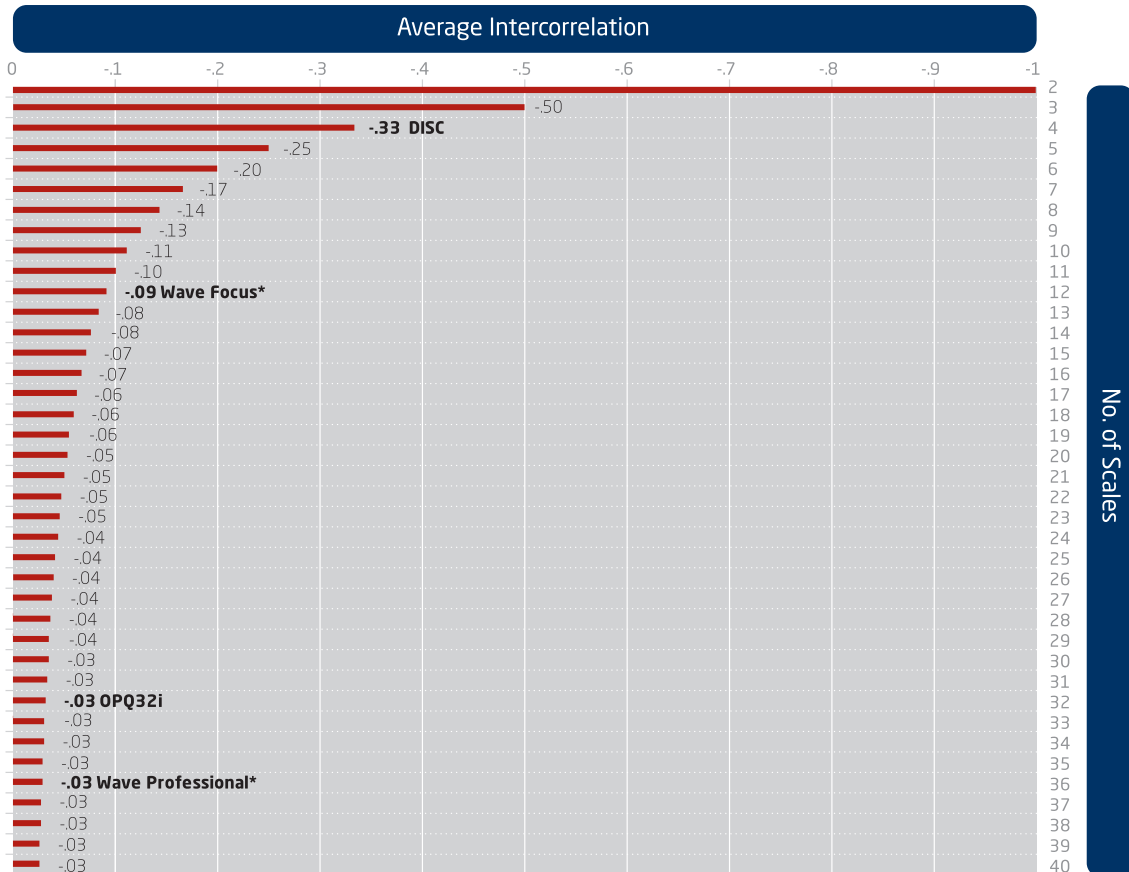> Analyzes information
> Ignores insults
> Organizes work

Polychotomous Most/Least

Simpler designs just ask Most (2 points) Least (0 points)

> Persuades others
> Analyzes information
> Ignores insults
> Organizes work

The important characteristic of ipsative scoring methods is that they create scores between scales that are interdependent. When scales are interdependent, it means that it is impossible to be high (or low) on every scale of the questionnaire. Also, the fewer primary scales assessed in the ipsative measure the greater the impact of the interdependency of the full scales themselves. This is shown in Graph 16.1.

Graph 16.1 The impact of the number of ipsative scales on the average intercorrelation among the scales.

Average Intercorrelation

No. of Scales	Value
2	
3	-.50
4	-.33 DISC
5	-.25
6	-.20
7	-.17
8	-.14
9	-.13
10	-.11
11	-.10
12	-.09 Wave Focus*
13	-.08
14	-.08
15	-.07
16	-.07
17	-.06
18	-.06
19	-.06
20	-.05
21	-.05
22	-.05
23	-.05
24	-.04
25	-.04
26	-.04
27	-.04
28	-.04
29	-.04
30	-.03
31	-.03
32	-.03 OPQ32i
33	-.03
34	-.03
35	-.03
36	-.03 Wave Professional*
37	-.03
38	-.03
39	-.03
40	-.03

DISC- Thomas International DISC – 4 scales- average intercorrelation of -.33

OPQ32 – Occupational Personality Questionnaire 32i – 32 scales – average intercorrelation of scales of - .03

Wave Professional has 36 scales in ipsative design giving average for the ipsative score component of – .03 and Wave Focus has 12 scales – giving average for this score component of - .09.

N.B. Wave Professional and Focus have a normative component so the actual average intercorrelations of combined scores are slightly positive rather than negative.

Ipsative questionnaires have the advantage that it is not possible to be high on every scale and with a higher number of scales, they present the relative preferences of the individual. By asking an individual to choose between different statements, we can measure that individual's relative preferences compared to simply asking the individual to describe him or herself (normative responses). A person may claim to enjoy all flavors of ice cream, but we can learn which flavor the person likes best by seeing which flavor they choose most often. Life in critical moments can be about making choices between a number of different options and ipsative scores are designed to better reflect these choices.

488

However, ipsative questionnaires also have limitations:

1. Potential for Increasing Error

One potential problem for ipsative scoring is that the questionnaire could present false or unrealistic choices to an individual. The individual in this case, could essentially be answering randomly without the responses being a representation of their underlying trait score. To reduce this it is important that the design of the ipsative questionnaire avoids presenting individuals with choices which are highly correlated while at the same time have similar mean endorsement values. This can also be a problem on normative forced-choice formats.

2. Limits Freedom/Acceptability

This is a property of forced-choice items in general that is not restricted solely to ipsatively scored items. The choice given may be restrictive and this may be irritating to the respondents where they feel that they are having to make a choice that does not adequately represent them. This may have the advantage of 'forcing out' some less socially desirable limitations or development needs, but at the cost of lowering the acceptability of the questionnaire to some users.

3. Interdependency

Digman (1997) provides evidence in favor of higher order factors of personality, where Alpha is Conscientiousness, Agreeableness and Emotional Stability and Beta is made up of Extraversion and Openness to Experience. Furthermore, Musek (2007) argues that Alpha and Beta are correlated, underpinning the 'Big One' factor of personality. If personality related characteristics are being measured, then the Big Five should be expected to show an overall small positive intercorrelation. Ipsative measures may therefore force the intercorrelations down unrealistically. The problem with interdependency is an increasing problem for questionnaires with few scales and where the true score intercorrelations are likely to be more strongly positive.

4. Norming and intra-individual vs. inter-individual comparisons

There is an argument which can be made that ipsative scores present a comparison within an individual, not between individuals, which makes standardizing against a comparison group inappropriate. The argument runs that a comparison of individual scores across individuals is not meaningful. The argument clearly holds more true where there are a smaller number of scales. However, as the number of scales increases and the degree of interdependency decreases (as shown in Graph 16.1), this issue diminishes until scale intercorrelations drop to near zero, and it is increasingly unlikely and unusual to find people who might be high on all the scales (particularly if there are relatively low intercorrelations between the scales).

5. Loss of a Degree of Freedom

If you have information on all but one of the scales, the final scale is just a linear composite of the other scales. The interdependency and this loss of a degree of freedom can affect some assumptions, some statistical analyses and critically,

489

some applications of the tool. To give an example of how this can become an issue, imagine that a user decides to use a tool with 16 ipsative scales designed to measure personality characteristics. Through job analysis 12 of these scales are identified to be job-related, and a linear composite is made by unit weighting and summing these 12 scales. This composite score, however, is also the negative or inverse composite of the four remaining scales (this creates an argument to also check that the four remaining scales not included in the equations are negative indicators of success!).

6.　Problems with Statistical Procedures

Some have argued that the problems of ipsative scoring are such that it is impossible to analyze or interpret using standard procedures (Hicks, 1970; Johnson, Wood & Blinkhorn, 1988) or that such scores can only be used in restricted contexts (Closs, 1996; Cornwell & Dunlap, 1994). However, others have argued that ipsative data is amenable to analysis using standard techniques and that its other properties often make it at least as useful as normative data (Gordon 1976; Saville & Wilson, 1991). Great care certainly needs to be taken with certain statistical procedures in interpreting the results. For example, there is no literature that the authors are aware of that canonical analysis is appropriate for ipsative data. Care also needs to be taken with factor analysis of ipsative only data. Bipolar factors often result in practice with one of the factors, for example, breaking across the opposite poles of the other factors making the results more difficult to interpret.

Despite these limitations, ipsative questionnaires can give a meaningful picture of the validity and utility with external criteria (that are not part of the score dependency). That is not to say that ipsative scoring cannot positively or negatively impact criterion validity, it can. For example, a scale like Agreeableness, when scored normatively, can have a positive correlation with an externally rated criterion such as overall performance, but when ipsatively scored can have a negative correlation depending upon the number of other scales and the intercorrelation among scales.

Ipsative and Normative Scoring

From a practical standpoint, it is useful to know both how positive someone is about themselves in general across the entire questionnaire (normative profile), and the relative emphasis they place on different behaviors (ipsative profile). However, asking for ipsative and normative scores on the same item is clearly lengthy, inefficient and potentially frustrating for respondents. The Ra-Ra (Rating-Ranking) method was developed to address this problem.

Saville Consulting Wave Dynamic Online Format

Saville Consulting Wave's dynamic online format is designed to be a more efficient method of creating a combined score with ipsative and normatively scored subcomponents. The Wave Styles questionnaire presents a page with six statements and a normative, free choice rating scale.

490

Professional Styles

Statements - Page 6 of 36

[?]

Please indicate to what extent you agree with the following statements.

	Very Strongly Disagree	Strongly Disagree	Disagree	Slightly Disagree	Unsure	Slightly Agree	Agree	Strongly Agree	Very Strongly Agree
I am **good at** dealing with uncertainty	○	○	○	○	○	○	○	○	●
I am **good at** seeing how things can be improved	○	○	○	○	○	○	○	○	●
I am **good at** understanding why **people do** things	○	○	○	○	●	○	○	○	○
People say **I make** a good first impression	○	○	○	○	○	●	○	○	○
I am **good at** negotiating	○	○	○	○	○	○	●	○	○
I am prepared to make big decisions	○	○	●	○	○	○	○	○	○

➡

© 2007 Saville Consulting. All rights reserved   Privacy   Accessibility

Once the ratings on the page are complete, normative ratings are saved and then ipsative ranks are calculated by the Oasys™ online assessment system. In this instance, it is easy to assign an ipsative score to most of the statements. Statements are rank ordered according to their normative ratings and an ipsative score is assigned. In the example, the ipsative score can be determined with the lowest ranked item of 'Disagree' being ranked last, 6th (and assigned an ipsative score of 1), the 'Unsure' rating being ranked 5th (and will receive an ipsative score of 2), 'Slightly Agree' will receive an ipsative score of '3', and 'Agree' and ipsative score of '4'. The two Very Strongly Agree statements are not possible to rank as they have equal ratings, so no ipsative score is assigned at this stage. If certain ranks cannot be determined from the normative ratings alone, a new page is created (see below). This page displays the tied items and asks the respondent to review the tied statements and use a forced-choice response format to determine which statement should be ranked higher.

Professional Styles

Statements - Page 6 of 36

[?]

Please select the statement that is most like you.

	Most
I am **good at** dealing with uncertainty	○
I am **good at** seeing how things can be improved	○

➡

© 2007 Saville Consulting. All rights reserved   Privacy   Accessibility

The item being assigned 'most' now results in a top ipsative score of '6' and the remaining item is assigned an ipsative score of '5'.

When more than two choices are ranked a most and least task is displayed. Where all the six statements have tied ratings, two 'most and least tasks' will be displayed one after another, assigning the most (6) and least (1) extreme rankings on the first ranking screen. On the subsequent screen the remaining four statements are displayed with the most receiving an ipsative score of five and the least receiving a score of two. Finally, the remaining two statements each receive a score of three and a half (there is no discernable increment in validity or reliability from seeking to differentiate the middle two statements). The questionnaire then presents a new page with six new statements.

By this mechanism, both ipsative and normative scores are computed efficiently and the two scores are also summed to create a combined score.

Benefits of the Dynamic Format:

1.   High Validity

In Project Epsom, the nine-item global Overall Performance scale had raw uncorrected validities of .32 normative, .34 ipsative and .36 for the Wave Professional Styles combined competency potential measures (N=280). This pattern replicates earlier criterion-related validity work on a sample of over 500 participants (MacIver et al, 2008). For Wave Focus, the pattern on Project Epsom is .24 ipsative, .27 normative and .29 combined (N=280). The effect sizes and sample size are insufficient to confirm the finding that one method is higher than the other at present, but the findings are consistent with the notion that the new format has at least equivalent (if not higher) validity than normative or ipsative counterpart scores alone. As further samples continue to build, this will be the subject of meta-analytic analyses.

We typically look to see if a questionnaire measures what it is designed to measure, but we can also look at the discriminant validity to see if it also measures competencies that it was not designed to measure. The average correlation with the 35 off-diagonals for the 36 Normative Competency Potential Wave Professional dimensions is .06 for normative, .02 for ipsative, and .03 for the combined score. The combined score having better (lower) discriminant validity than normative and practically on a par with ipsative discriminant validity. These differences are particularly important where differential prediction is needed across different possible outcomes, such as prioritizing individual development actions or career guidance.

2.   High Reliability

Support for the use of the new combined method with alternate forms for normative, ipsative and combined styles scores all averaging in the .80's. Refer to the Reliability chapter for further information.

3.   High Variance

The method allows for the creation of normed scores at the facet level. The ipsative and normative scores are combined, and the use of the nine-point agreement scale with blocks of six statements significantly increases the number of raw score points available per item. This ensures sufficient variance at the facet level thereby allowing for shorter scales to be scored and profiled.

4.   Stronger Intra-individual Differentiation

Another consideration is whether the profile of scores provided by a Wave Styles questionnaire gives better intra-individual variation. This, in combination with good convergent and discriminant validity, will lead to better validity in making differential selection decisions (between different jobs), or career planning and identifying individual development actions.

To investigate this an analysis was conducted. For each respondent in the Wave standardization sample a standard deviation was calculated across the twelve competency potential section scores generated using each of the different scoring methods. This provided an SD for each individual based on how differentiated the respondent's score was for each of the scoring methods. The average of this SD was calculated across the 1,153 participants in the study.

The differentiation across the twelve sections for the Wave Competency Potential scores for normative is 1.26, for ipsative 1.93, and for the combined score is 1.81.

The results indicate that ipsative has the greatest intra-individual differentiation, but that the combined score was very close to this.

5. Candidate Acceptability

Using a combined method results in a lower proportion of forced-choice questions being asked than with a forced-choice fully ipsative design (such as Thomas International DISC or OPQ32i). This has the benefit of increasing the acceptability of the user experience. The dynamic format of Wave combines the candidate-friendly free choice rating format with the more discerning ranking format in order to maximize usability and efficiency for the respondent.

6. Making Faking More Complex

No questionnaire is unfakable, but the addition of the ranking component increases the complexity that would be required in the process of faking. An individual has to try to control a greater number of variables when doing the rating and ranking tasks.

7. Making Distortion Easier to Detect

A combined score is always plotted, but when a difference of three or more stens between the ipsative score and the normative score exists this difference is plotted on the profile chart. The user can see which response format resulted in the higher score and by how much. The reliability and validity indicate that the normative, ipsative and combined scores are all valid, so it provides an opportunity for the individual interpreter to find out where distortion may be taking place and find out the reasons for the difference between the two scores. It could be that the ipsative score on a scale is a better representation of behavior than the normative, or vice versa.

493

Limitations of the Dynamic Format

No format results in a perfect, error free representation of trait scores. While the new format conveys clear advantages by dynamically presenting ratings and rankings, not all of the limitations present in normative and ipsative scores are eradicated. For instance, the interactive nature may result in people altering their rating responses to avoid ties as this will reduce the number of ranking tasks they complete to break the ties (and finish the questionnaire sooner). This is something to be aware of when interpreting the profile, particularly with respondents who were not very motivated to complete the questionnaire (see Feedback chapter). Despite this, the results discussed above indicate that the dynamic online format has advantages over ipsative and normative scores alone.

Response Format Summary

The new response format of Saville Consulting Wave provides an efficient mechanism that increases the fidelity of measurement, allowing more information to be measured in less time. The validity and reliability is equal to if not better than either the normative and ipsative components in the score (and time will tell if the combination has better convergent validity still than either of the individual components). This is accomplished while reducing the positive intercorrelation present within normative only scores of the Wave Professional Styles intercorrelation matrix and allowing ipsative and normative differences within the profile to pinpoint potential areas of distortion.