

Introducing Item Pool Visualization: A method for investigation of concepts in self-reports and psychometric tests

Methodological Innovations
September-December 2019: 1–10
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2059799119884283
journals.sagepub.com/home/mio


Michael Dantlgraber¹ , Stefan Stieger^{1,2} and Ulf-Dietrich Reips¹

Abstract

In this article, we introduce Item Pool Visualization. It is an illustration system that locates items and item pools (scales) from multiple psychological instruments regarding their commonality and distinguishability along several dimensions of nested radar charts. The application of Item Pool Visualization creates illustrations that represent different item pools by different circles that do not overlap. Item Pool Visualization illustrates a comparison of different structural equation models that are estimated with the same data. It combines the advantages of general and correlated factor models when evaluating psychological instruments. Furthermore, in contrast to other visualization methods, Item Pool Visualization provides an empirically driven categorization of psychological constructs and their subconcepts (facets) that is suited to provide professionals with help in comparing psychometric constructs, questionnaires, and selecting tests.

Keywords

Item Pool Visualization, construct validation, scale comparison, validity, structural equation model, factor analysis, factor structure, visualization, nomological network

Introduction

In both applied and research settings, psychologists often face the challenge of selecting the right questionnaires or tests suited to help answering their specific questions. For example, if one wants to use a questionnaire to measure individual differences in personality, one has to choose between several well-developed personality scales (i.e. questionnaires) that all deliver reliable results. These scales may, however, focus on different aspects of the same psychological construct even if their labels are very similar or identical (e.g. different *Extraversion Scales* focus on extraverted behavior in different situations). Thus, how to make an informed selection?

In this article, we introduce Item Pool Visualization (IPV) and show how it facilitates scale comparisons of clearly defined item pools (e.g. published scales). This is relevant in applied research and also for theory development, that is, to understand how well certain instruments capture certain constructs. We argue, for example, that two questionnaires that are both aiming to measure *Self-Esteem* can have a different understanding of the same concept. While one questionnaire

primarily evaluates the level of *Self-Esteem* regarding the trust in *Social and Task-Related Abilities*, the other primarily considers the lack of self-doubts (lack of *Negative Self-Esteem*).

IPV illustrates comparisons of factor loadings when different structural equation models (SEMs; Kaplan, 2001; Wright, 1921) or rather factors and further specified factors are estimated with the same data. With it we seek to provide an illustration system that locates item pools (scales) regarding their commonality and distinguishability along several dimensions of nested radar charts. IPV creates clear illustrations to represent different item pools by different circles that do not overlap. For this reason and others, IPV substantially differs from traditional Venn diagrams (Venn, 1980), which also illustrate relations between variables depicted as circles.

¹Department of Psychology, University of Konstanz, Konstanz, Germany

²Department of Psychology, University of Vienna, Vienna, Austria

Corresponding author:

Michael Dantlgraber, Department of Psychology, University of Konstanz, Box 31, 78457 Konstanz, Germany.

Email: michael.dantlgraber@uni-konstanz.de



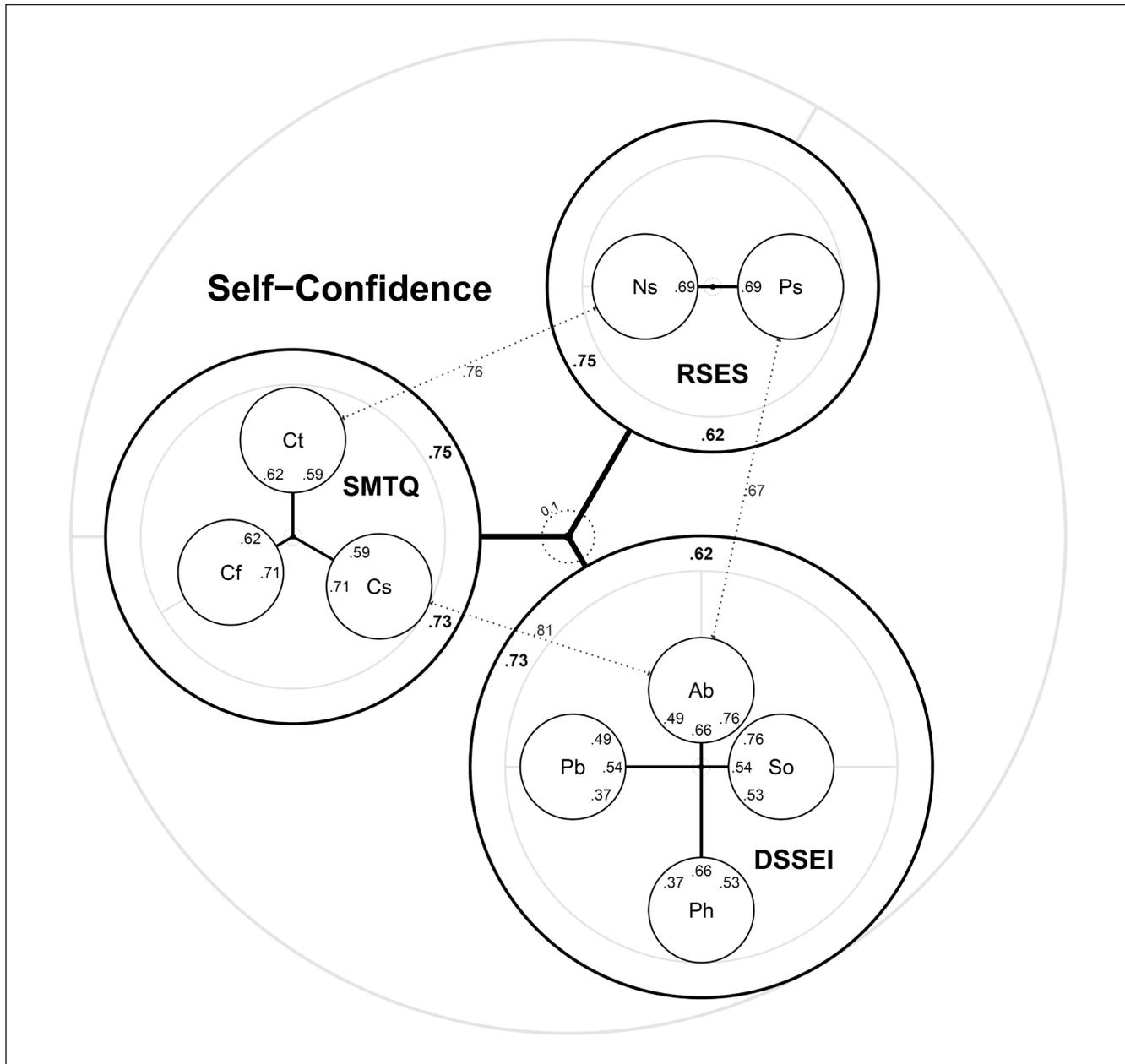


Figure 1. Nested IPV of the DSSEI, SMTQ, and RSES. Circles represent item pools. Numbers within circles represent latent correlations between different item pools.

So: Social Competence; Ab: Task-Related Abilities; Ph: Physical Appeal; Pb: Public Presentation; Cf: Confidence; Cs: Constancy; Ct: Control; Ns: Negative Self-Esteem (recoded); Ps: Positive Self-Esteem.

Figure 1 shows a nested IPV of three different multi-facet questionnaires that measure positive self-concepts, namely, the Domain-Specific Self-Esteem Inventory (DSSEI; Hoyle, 1991), the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965), and the Sports Mental Toughness Questionnaire (SMTQ; Sheard et al., 2009).

The goal of this study is to present and discuss IPV as a methodological innovation that will help in understanding psychological constructs and in the analysis and selection of psychological instruments. In this article, we explain the

structure of IPV and its interpretation on the basis of the example shown in Figure 1. We discuss the advantages of IPV over traditional SEM illustrations. Because IPV illustrates a model comparison, it combines different benefits of different estimations in one single illustration. As shown in Figure 1, IPV illustrates comparisons of facets within and between questionnaires like illustrations of correlated factor models do. But it also illustrates superordinate commonalities like illustrations of general or a hierarchical factor models do. The combination of different information in one

single illustration enables the discovery of additional similarities of and differences between psychological instruments that are easily overseen in traditional scale comparisons. This aspect is described in detail in the following sections.

Furthermore, IPV not only illustrates more information than traditional visualization methods but also illustrates the information more meaningful. Because of the use of nested radar charts, the arrangement of the included questionnaires and facets (i.e. larger and smaller defined item pools) carries a meaning. This is not given in any of the traditional SEM illustrations. IPV's arrangement is based on clear rules that define the distances of subpools to the center of a radar chart, representing a superordinate item pool. In the next sections, we first systematically explain the estimation of these *center distances*, which constitute the basis of IPV.

Method and Results

Materials

To illustrate IPV, we use a large dataset containing three different questionnaires that measure positive self-concepts, namely the DSSEI (Hoyle, 1991), the RSES (Rosenberg, 1965), and the SMTQ (Sheard et al., 2009).

Sample and procedure

The sample was drawn in Germany and Austria and consists of 2272 German-speaking participants who filled out German versions of the DSSEI, the RSES, and the SMTQ. The sample constitutes a community-based sample and is age-stratified. Of the participants, 1265 (56%) reported to be female, 980 (43%) reported to be male, 27 (1%) did not state their sex. Participants' mean reported age was 39.8 years (standard deviation = 17.7). Regarding highest educational level, 547 (24%) had a university degree, 783 (34%) had high school graduation, 512 (23%) had an apprenticeship certificate, 333 (15%) completed secondary education, and 86 (4%) had no degree.

Analysis

To explain the general structure of IPV, we first present an IPV of the DSSEI and its four facets (*Social Competence*, *Task-Related Abilities*, *Public Presentation*, and *Physical Appeal*). This implies estimations of two SEMs and subsequent calculations.

In the next step, we estimate and present a nested IPV, where the DSSEI and its facets are compared with the RSES and the SMTQ and their facets (see Figure 1). This implies estimations of three SEMs and subsequent calculations.

IPV is suited to do a comparison of different SEMs using the same data. Uniquely, in comparison with other methods, IPV illustrates in a radar chart the proportional rise of squared factor loadings when an overall item pool (e.g. the whole questionnaire) is split into some smaller and further

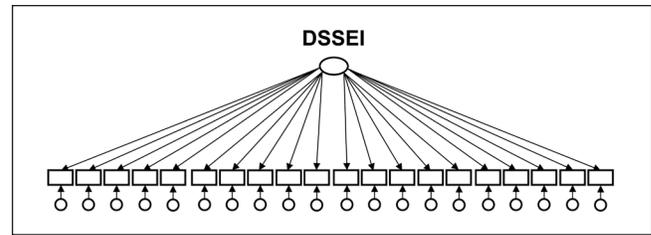


Figure 2. General factor model of the Domain-Specific Self-Esteem Inventory (DSSEI).

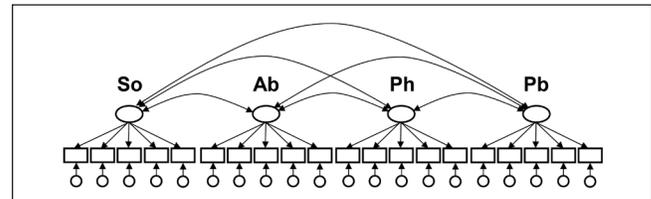


Figure 3. Correlated factor model of the Domain-Specific Self-Esteem Inventory (DSSEI).

So: *Social Competence*; Ab: *Task-Related Abilities*; Ph: *Physical Appeal*; Pb: *Public Presentation*.

specified subpools (e.g. facets). To provide an example, we first show this for the DSSEI only.

The first step when generating an IPV always is the estimation of a general factor model. There, a single factor is extracted from an overall item pool including all items, in this example, all items of the DSSEI. Figure 2 shows a traditional illustration of this model.

As a second step, a correlated factor model is estimated where more factors are extracted from smaller and more specific subpools. In this example, four correlated factors are extracted from four smaller item pools representing the four facets of the DSSEI as described by Hoyle (1991; *Social Competence*, *Task-Related Abilities*, *Public Presentation*, *Physical Appeal*). Figure 3 shows a traditional illustration of the second model.

Note that SEM is needed when estimating a correlated factor model, that is, it cannot be replaced by a traditional factor analysis with a fixed number of correlated factors. This is because only SEM allows a clear allocation of the items to the factors.

When estimating the two models (Figures 2 and 3), different factor loadings are estimated for the same items (see Table 1). This is because the general factor is extracted from a common item pool (Figure 2) and the correlated factors are extracted from further specified item pools (Figure 3). Table 1 shows that the factor loadings of the correlated factor model (Figure 3) tend to be higher than the loadings of the general factor model (Figure 2). This is due to the correlated factor model representing the facets (i.e. normally more similar items) that can be better represented by single factors.

Table 1. Basic calculation of the IPV for the DSSEI.

Fc	Item	Factor loadings		Ratio of squared loadings	Center distance
		General factor model	Correlated factor model		
So	I usually feel as if I have handled myself well at social gatherings.	.58	.73	1.58	0.58
	I feel secure in social situations.	.69	.83	1.45	0.45
	I feel confident of my social behavior.	.66	.79	1.43	0.43
	I am often troubled with shyness. (R)	.40	.31	0.60	<.00
	At social gatherings, I am often withdrawn, not at all outgoing. (R)	.45	.43	0.91	<.00
Ab	I feel as if I lack the necessary skills to really succeed at the work I do. (R)	.27	.30	1.23	0.23
	I am able to do things as well as most other people.	.60	.66	1.21	0.21
	I usually expect to succeed at the things I do.	.52	.63	1.47	0.47
	I almost always accomplish the goals I set for myself.	.57	.69	1.47	0.47
	In general, I feel confident about my abilities.	.64	.73	1.30	0.30
Ph	I feel that others would consider me to be attractive.	.56	.80	2.04	1.04
	I'm not as nice looking as most people. (R)	.21	.37	3.10	2.10
	I feel confident that my physical appearance is appealing to others.	.57	.83	2.12	1.12
	I am satisfied with the way I look.	.56	.71	1.61	0.61
	I feel unattractive compared to most people my age. (R)	.45	.57	1.60	0.60
Pb	When I speak in a large group discussion, I usually feel sure of myself.	.67	.87	1.69	0.69
	I enjoy being in front of large audiences.	.57	.75	1.73	0.73
	I feel quite confident when speaking before a group of my peers.	.66	.80	1.47	0.47
	I find it very hard to talk in front of a group. (R)	.45	.68	2.28	1.28
	When I talk in front of a group of people my own age, I am usually somewhat worried or afraid. (R)	.53	.67	1.60	0.60

IPV: Item Pool Visualization; DSSEI: Domain-Specific Self-Esteem Inventory; Fc: Facet; So: *Social Competence*; Ab: *Task-Related Abilities*; Ph: *Physical Appeal*; Pb: *Public Presentation*.

Items followed by (R) were reverse-scored before analysis.

Calculating center distances

In order to quantify the increase of the loadings when switching between the two models (Figures 2 and 3), IPV uses the ratios of the squared loadings. More precisely, the squared loadings of the correlated factor model are divided by the squared loadings of the general factor model (see Table 1).

A ratio of squared loadings equal to 1 signifies that the estimated loadings of the respective item are identical in the correlated and in the general factor model. Furthermore, it signifies that the factor extracted from the smaller subpool does not explain more variance of the respective item than the factor extracted from the larger overall item pool. In other words, a ratio of squared loadings equal to 1 represents no difference. For this reason, in a basic IPV calculation, 1 is subtracted from each ratio to represent a missing difference with 0. The results of these calculations are called *center distances*.

Center distances represent the proportional increase of the explained item variance, if the items are allocated to smaller subpools (Figure 3), compared to a larger common pool (Figure 2). A center distance of 0 represents no increase. By definition, no increase includes a possible decrease of

explained variance, that is, negative values are treated as 0 (see Table 1).

Visualizing items of the DSSEI

Center distances are used for locating the items along several facet dimensions in a radar chart (see Figure 4).

In an IPV, the center of the radar chart represents the item variance that is explained by the factor extracted from this overall item pool. Therefore, center distances illustrate how much each item is more strongly associated with a smaller subpool in comparison with the overall item pool. For example, the first item of the facet *Physical Appeal* (*I feel that others would consider me to be attractive*) has a center distance of 1.04 (see Table 1). This implies a 104% increase of the explained variance of this item if it is viewed as a *Physical Appeal* item instead of just as a *Self-Esteem* item.

Visualizing facets of the DSSEI

Items that are located on the same facet dimension in Figure 4 can build an item pool representing the respective facet. In

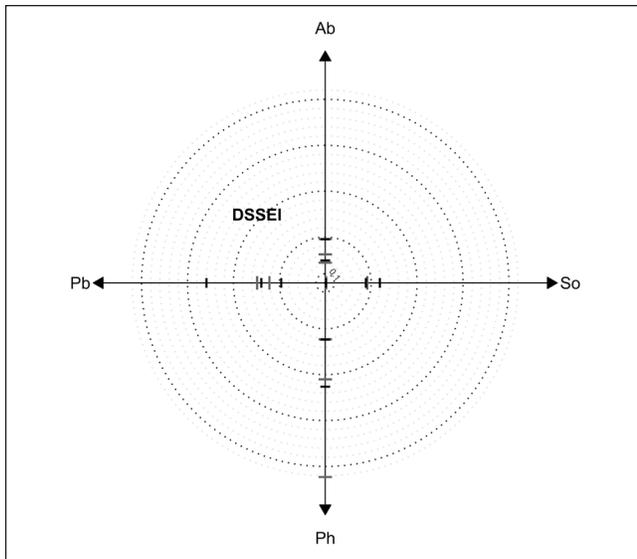


Figure 4. Item location of the DSSEI on facet dimensions when using IPV.

So: Social Competence; Ab: Task-Related Abilities; Ph: Physical Appeal; Pb: Public Presentation.

The dotted circles are the grid, showing axis scaling. For clearer distinction, every second item is illustrated in different lengths and color.

Figure 5, the items build four item pools representing the four original facets of the DSSEI (*Social Competence, Task-Related Abilities, Public Presentation, and Physical Appeal*).

The four circles representing the item pools of the four facets are located by using the *mean center distance* of the included items. The mean center distance is illustrated as the distance from the center of the radar chart to the circle edge indicated by the thick black radial lines. IPV uses this illustration system to avoid cluttered figures that would come from overlapping circles, numbers, and labels. Non-overlapping figures ensure clear communication of meaning and avoid misunderstandings that could, for example, come from erroneous attribution of labels to the wrong component. Note that with this illustration system, it is always possible to adjust the scaling of the specific dimensions so that the circles do not overlap.

The numbers within a circle represent the latent correlations between this item pool (facet) and the other item pools (facets). They are estimated with the model shown in Figure 3. Correlations are arranged clockwise in the same order as the facets. The IPV of the DSSEI shows that the facets are not balanced (see Figure 5). Contents of facets near the center tend to be better represented by the overall questionnaire, because they tend to be captured well by the general factor, that is, have more commonalities with contents of other facets.

Figure 5 shows that the two facets that are nearest to the center (*Social Competence, Task-Related Abilities*) are correlated the most with other facets. That implies that participants with high DSSEI scores often show a response pattern that particularly consists of high ratings in these two facets.

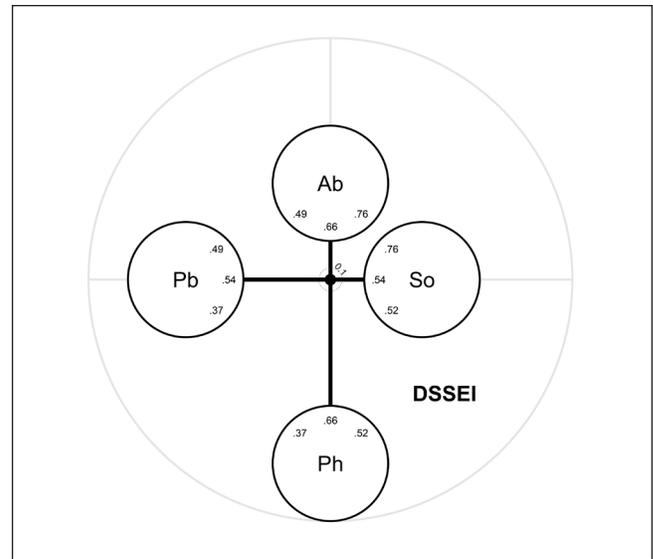


Figure 5. Facet location of the DSSEI when using IPV.

So: Social Competence; Ab: Task-Related Abilities; Ph: Physical Appeal; Pb: Public Presentation.

Requirements for IPV

IPV has three requirements. (1) The data must allow SEM (see Kaplan, 2001). (2) All estimated factor loadings in each SEM should be positive. However, if all loadings of the same item pool are negative, it is possible to recode the items (see section “Visualizing a comparison of the DSSEI with the RSES and the SMTQ”). (3) All estimated factor loadings in each SEM should not be too low. We recommend to exclude items that have factor loadings below .1, which means that less than 1% of their variance can be explained by the respective factor (not relevant in the examples presented in this article).

Visualizing a comparison of the DSSEI with the RSES and the SMTQ

As demonstrated in the previous sections, IPV is a nested system that compares factor loadings when an item pool is split into smaller subpools. This system allows item pool comparisons on several hierarchical levels. Subpools (facets) can contain further subpools and item pools (questionnaires) can also be combined with other item pools (questionnaires) and form a superordinate item pool. The latter is needed to compare the DSSEI with the RSES and the SMTQ that were also created to measure a positive self-concept. Such a comparison implies three hierarchical levels. Therefore, three SEMs are estimated (see Figures 6 to 8).

In the first SEM (Figure 6), one factor is extracted from a superordinate item pool (from all items of the three questionnaires). The item *In general, I feel confident about my abilities* from the DSSEI facet *Task-Related Abilities* and the item *I feel confident about my social behavior* from the DSSEI facet *Social Competence* have the highest loadings on the

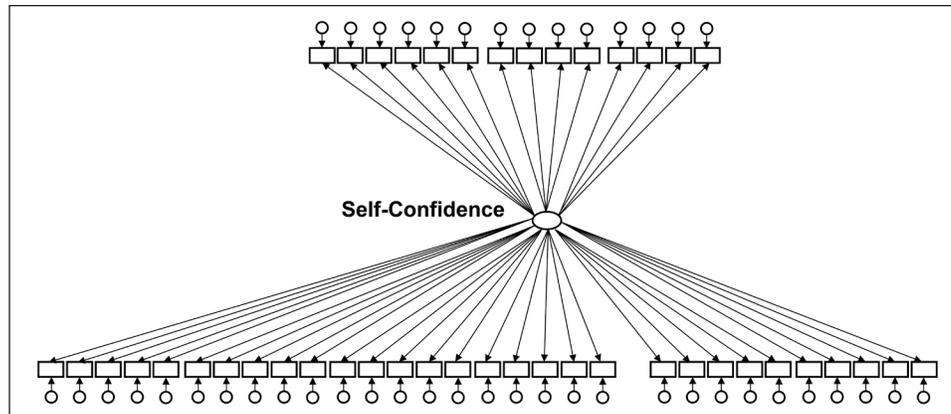


Figure 6. General factor model including all items of the DSSEI, SMTQ, and RSES.

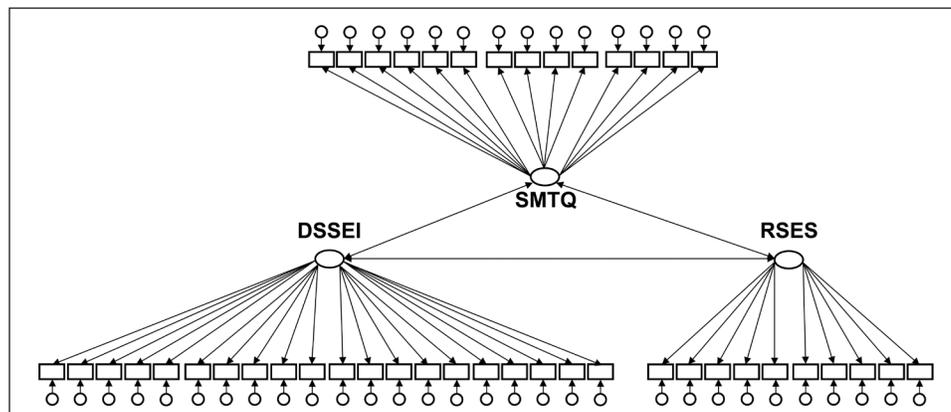


Figure 7. Correlated factor model of the main factors of the DSSEI, SMTQ, and RSES.

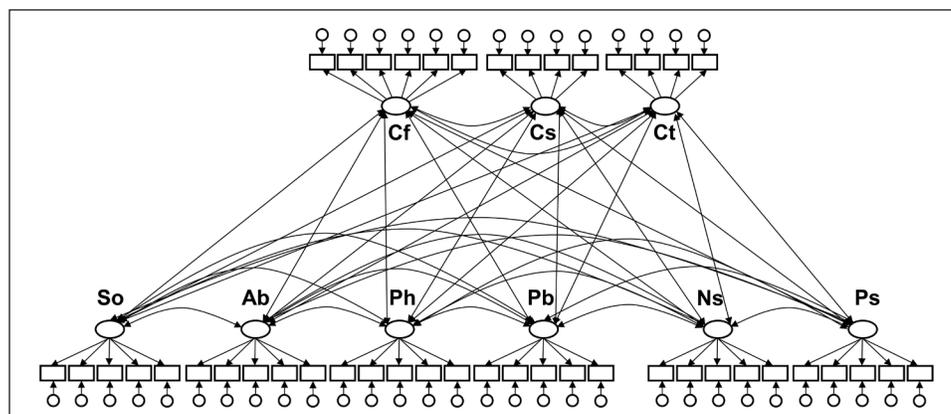


Figure 8. Correlated factor model of the facets of the DSSEI, SMTQ, and RSES.

So: Social Competence; Ab: Task-Related Abilities; Ph: Physical Appeal; Pb: Public Presentation; Cf: Confidence; Cs: Constancy; Ct: Control; Ns: Negative Self-Esteem (recoded); Ps: Positive Self-Esteem.

superordinate general factor, including all items of the three questionnaires ($r = .66$, $r = .62$). For this reason, we named the superordinate factor (Figure 6), *Self-Confidence*.

In the second SEM (Figure 7), three correlated factors are extracted from three subpools (i.e. three sections of the

overall item pool) representing the three questionnaires (DSSEI, RSES, SMTQ).

In the third SEM (Figure 8), nine correlated factors are extracted from nine subpools representing the facets of the three questionnaires. Within the DSSEI and the SMTQ, we

grouped items to facets as indicated by the scales' original authors. Within the RSES, we grouped items into two subscales as suggested in the past by some researchers (*Positive vs. Negative Self-Esteem*; for a recent discussion, see Supple et al., 2013). In order to meet the requirements of an IPV (see section "Requirements for IPV"), all items of the facet *Negative Self-Esteem* were recoded.

Due to the three model estimations (Figures 6 to 8), three different factor loadings are estimated for each item. Therefore, two center distances can be calculated for each item using the same procedure as described above. One center distance can be calculated that represents the comparison of the first and the second SEM (Figures 6 and 7), that is, the comparison of the three tests in a superordinate radar chart. In addition, one center distance can be calculated that represents the comparison of the second and the third SEM (Figures 7 and 8), that is, the comparison of the facets within the three subordinate radar charts. When using *mean center distances*, a nested IPV can be drawn that includes information of the three SEMs (previously shown in Figure 1).

Figure 1 shows the relationship in meaning of the DSSEI, the SMTQ, and the RSES using IPV. The correlation numbers in bold represent the latent correlations of the main factors of the three questionnaires (estimated by the model shown in Figure 7). Note that the center distances within the three questionnaires are drawn on the same scale, but the center distances between the three questionnaires are triple that in order to avoid overlapping circles. In Figure 1, this is indicated by the dotted lines, each representing a center distance of 0.1.

Interpretation of the center distances in Figure 1

Figure 1 integrates three nested radar charts (representing the three questionnaires) in one superordinate radar chart. Like every radar chart, IPV is interpreted by analyzing distances along the dimensions. We recommend starting the interpretation of an IPV from the center of the superordinate radar chart (*Self-Confidence*, see Figure 1). The center of the superordinate radar chart has the shortest distance to the edge of the overall circle of the DSSEI, meaning the general factor of the DSSEI differentiates least from the superordinate factor (*Self-Confidence*) followed by SMTQ and RSES.

Next, we interpret the facet level. The center of the DSSEI has the shortest distance to its facets *Social Competence* and *Task-Related Abilities*. That means that these facets differentiate least from the general factor of the DSSEI, which, as described above, differentiates least from the superordinate factor (*Self-Confidence*). This implies that *Self-Confidence*, *Social Competence*, and *Task-Related Abilities* are to some degree interchangeable terms because they contain little unique variance.

In some cases, facets between different questionnaires are correlated more strongly than the general factors of the respective questionnaires. This can be viewed as an indicator

that the specific facet dimensions of the different questionnaires have something in common. In an IPV, facet correlations between different questionnaires that are higher than the correlations of their general factors are drawn with dotted arrows. Figure 1 shows, for example, that the specific dimension *Task-Related Abilities* of the DSSEI has something in common with the specific dimensions *Constancy* and *Positive Self-Esteem* of the SMTQ and the RSES.

Interpretation of the factors in Figure 1

Using IPV facilitates the interpretation of factors. In Figure 1, the general factor of the RSES (representing *Self-Esteem*) shows a higher correlation with the general factor of the SMTQ (representing *Mental Toughness*) than with the general factor of the DSSEI (also representing *Self-Esteem*). This is surely not an expectable result because both the RSES and the DSSEI share the same label, that is, are indicated to measure *Self-Esteem*.

However, Figure 1 also shows that the facet *Negative Self-Esteem (recoded)* of the RSES differentiates least from its general factor (*Negative Self-Esteem (recoded)* is nearer to the center than *Positive Self-Esteem*). Furthermore, the facet *Negative Self-Esteem (recoded)* of the RSES shows a high correlation with the facet *Control* of the SMTQ, that is, not having *Negative Self-Esteem* has something in common with self-control. This commonality can explain the unexpectedly high correlation of the general factors of the RSES and the SMTQ that both include these contents.

In conclusion, Figure 1 shows that the overall item pool focuses more on contents like task related and social skills than on contents like self-control or physical attraction. Because these contents can be found in all questionnaires, they might be more important for a positive self-concept.

Even if IPV cannot absolutely locate item pools semantically, interpreting their relations facilitates the understanding of the questionnaires and their operational definition. Consequently, IPV enhances the understanding what questionnaires really measure. In this and the previous section, the interpretation of Figure 1 was explained. In the next section, general interpretation rules for IPV are listed.

General interpretation rules for IPV

1. In an IPV, circles represent item pools.
2. The only distances that have a meaning in an IPV are *center distances*, that is, distances from the center of a certain radar chart to edges of other circles or items.
3. The center distance of a single item represents the proportional increase of its explained variance when it is allocated to a smaller item pool instead of a larger reference pool. For example, suppose the *squared loading* of an item increases from .2 in the general factor model to .3 in a correlated factor model. In this case, the increase of a squared loading from .2 to .3

would be illustrated with a center distance of 0.5, implying an increase of 50%.

4. The center distance of a circle (item pool) represents the mean center distance of its composite items.
5. It is possible to define a route of center distances through several radar charts of different hierarchical levels. The route starts at the center of the most superordinate radar chart and ends at the circle edge of a subordinate radar chart. In the center of this subordinate radar chart, the route can be continued.
6. The numbers within the circles represent the latent correlations of factors extracted from different item pools when using SEM.
7. Dotted arrows represent the latent correlation between subordinate circles (item pools) that do not share the identical superordinate circles (item pools), that is, they illustrate the latent correlation between facets of different questionnaires. Unless otherwise specified, dotted arrows are used when the latent correlations of the facets are higher than the latent correlations of the questionnaires, implying an additional similarity between these facets. The length of the arrows does not have any meaning.

Discussion

Why using IPV?

IPV focuses on content-related differences of item pools. It can be used to illustrate the facet structure of a single questionnaire or to compare different questionnaires regarding their facet structures. Illustrating the facet structure of a single questionnaire primarily serves the purpose to show the specificity or the balancing of the facets. In most cases, facets are not balanced, which should be taken into account when interpreting results from data collection with an assessment scale. For example, when interpreting a possible correlation between the overall DSSEI score and another variable, it should be taken into account that the overall DSSEI score represents *Social Competence* and *Task-Related Abilities* the most. In other words, IPV illustrates that the DSSEI predominantly operationalizes *Self-Esteem* with *Social Competence* and *Task-Related Abilities*. In this way, IPV differentiates the DSSEI from the RSES that predominantly operationalizes *Self-Esteem* with *Negative Self-Esteem (recoded)* and therefore offers an opportunity for quick comparisons.

Furthermore, consisting of a nested system of item pool labels, IPV offers the opportunity to systematically investigate the meaning of assessment scales. This is an important issue because authors rarely name their scales on the basis of strict empirical rules (e.g. following the content of the items showing the highest factor loadings). Normally, authors name their scales after an existing construct or theory irrespective of whether the item pools (scales) are really

representing them in the end of the development process (e.g. Bech et al., 2003; Oviedo-García, 2007).

For example, a scale that exclusively consists of items related to conspicuous party behavior is typically named as an *Extraversion Scale* (and not as a *Conspicuous Party Behavior Scale*). This procedure is obviously problematic. If other authors use this scale later in their studies, in future literature their findings will be related to *Extraversion* and not to *Conspicuous Party Behavior*.

It is important to note that the above example describes a restriction of content validity that cannot be detected with measures of reliability or by evaluating fit indices as criteria (Hu and Bentler, 1999). Naturally, a *Conspicuous Party Behavior Scale* can be very reliable and can show very good fit indices (e.g. in a general factor model). Moreover, the content restriction of this *Extraversion Scale* consisting of some very similar items could be the reason for its high reliability and its good fit indices. Even if low reliability reduces validity, high reliability does not guarantee validity. IPV was developed to facilitate content-related scale comparisons, which cannot be done by comparing measures of reliability or fit indices. In practice, scales are often selected on the basis of reliability measures and fit indices only. IPV is a tool to evaluate if the content of a scale fits its purpose.

Visualizing the Big Five personality traits

Some readers may wonder how an IPV of the Big Five personality traits would look like. Naturally, it would be possible to separately illustrate each of the five dimensions with their respective facets. A further possibility is based on the fact that the Big Five personality traits are not uncorrelated, as was originally assumed (see Van der Linden et al., 2010). For this reason, it would be possible to illustrate the dimensions pairwise as in Hofstee et al. (1992) with their AB5C model. Researchers who are interested in this approach should keep in mind that probably not all possible pairs match the requirement for IPV that the loadings of the factors (in particular of the superordinate factor) should not be too low (see section "Requirements for IPV"). However, it seems likely that the three substantially positively correlated dimensions *Emotional Stability* (neuroticism recoded), *Agreeableness*, and *Conscientiousness* (see Van der Linden et al., 2010) could be illustrated in one single figure, at least with some smaller item adjustments. An IPV of these three dimensions would look similar to Figure 1.

Relation to higher order factor models

IPV is based on comparisons of different SEMs that represent different hierarchical levels (see section "Method and Results"). In this regard, some may ask why not estimate one single higher order factor model. For this reason, Figure 9 shows how a higher order factor model would look like that is comparable with the IPV shown in Figure 1:

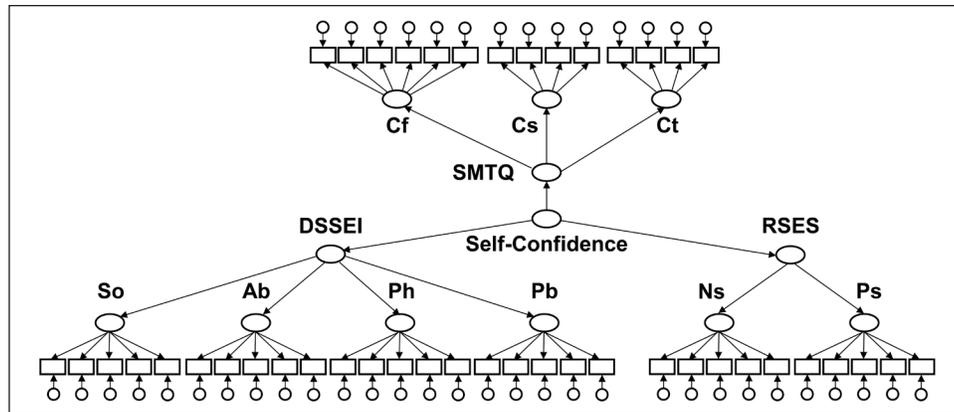


Figure 9. Higher order factor model of the DSSEI, SMTQ, and RSES.

So: Social Competence; Ab: Task-Related Abilities; Ph: Physical Appeal; Pb: Public Presentation; Cf: Confidence; Cs: Constancy; Ct: Control; Ns: Negative Self-Esteem (recoded); Ps: Positive Self-Esteem.

- In a higher order factor model, the arrangement of the questionnaires and facets has no meaning. In IPV it does.
- Because of using nested circles, IPV allows zooming in and out of the figure and therefore can be more easily extended.
- IPV more strongly emphasizes that factors are based on item pools and that the validity of the estimated factors depends on the validity of the item pools.
- Most important: Higher order factor models offer no opportunity to directly compare facets within or between questionnaires. That means when interpreting a higher order factor model, it cannot be easily detected if facets of different questionnaires having the same labels substantially differ or if facets of different questionnaires having different labels are in fact very similar (as with the facets *Task-Related Abilities* (Ab) and *Constancy* (Cs) in the discussed example). For direct facet comparisons, a correlated factor model is needed (see Figure 8), which is, however, not hierarchical or nested. IPV (see Figure 1) combines advantages of higher order and correlated factor models.

Limitations of IPV

IPV was developed to facilitate comparisons of similar and therefore positively correlated scales. Even if negative correlated scales can be recoded in some cases (see section “Requirements for IPV”), IPV cannot be used to illustrate a complex network of both positively and negatively correlated scales.

Furthermore, the hierarchical structure of IPV and its zoom feature potentially offers the opportunity to extend the illustration with further positively correlating scales building a large network of constructs. However, when using the procedure described above, the potential to create a large network of constructs is limited by the reasonable length of a

study. Naturally, participants may not fill in more than a certain number of scales in one single study (this issue concerns all types of modeling).

We think this problem can be solved in IPV when adapting the procedure and combining different samples that focus on different hierarchical levels. While one sample could focus on the internal structure of a single questionnaire (using all items), another sample could focus on the connection of different questionnaires (using well-selected short forms representing identical factors). This would limit the lengths of each study. Nevertheless, even if this procedure seems reasonable to us, its practicability is not proven yet.

Conclusion

In this article, we introduced IPV that offers the opportunity to locate items and questionnaires in large networks and to zoom in and out of these networks. IPV combines benefits of higher order and correlated factor models within one single illustration. This enables to discover additional similarities and differences of assessment scales that are easily overseen with traditional visualizations. Furthermore, in contrast to other visualization methods, IPV provides an empirically driven categorization of questionnaires and their facets that is suited to provide professionals with help in comparing constructs, questionnaires, and selecting tests.

Acknowledgements

The authors would like to thank Nils Petras for creating the figures of the IPV and for working on an R software package for IPV. Furthermore, the authors would also like to thank Tim Kuhlmann, Nils Petras, and Lars Hilbert for their helpful suggestions and comments.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by Ulf-Dietrich Reips' chair in the Department of Psychology of the University of Konstanz.

ORCID iD

Michael Dantlgraber  <https://orcid.org/0000-0001-5386-1259>

Data accessibility statement

All materials are available on request from the corresponding author (michael.dantlgraber@uni-konstanz.de).

References

- Bech P, Olsen LR, Kjoller M, et al. (2003) Measuring well-being rather than the absence of distress symptoms: A comparison of the SF-36 Mental Health subscale and the WHO-Five Well-Being Scale. *International Journal of Methods in Psychiatric Research* 12: 85–91.
- Hofstee WKB, De Raad B and Goldberg LR (1992) Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology* 63: 146–163.
- Hoyle RH (1991) Evaluating measurement models in clinical research: Covariance structure analysis of latent variable models of self-conception. *Journal of Consulting and Clinical Psychology* 59: 67–76.
- Hu L and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6: 1–55.
- Kaplan D (2001) *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: SAGE.
- Oviedo-García M (2007) Internal validation of a biodata extraversion scale for salespeople. *Social Behavior and Personality* 35(5): 675–692.
- Rosenberg M (1965) *Society and the Adolescent Self-Image*. Princeton NJ: Princeton University Press.
- Sheard M, Golby J and Van Wersch A (2009) Progress toward construct validation of the Sports Mental Toughness Questionnaire (SMTQ). *European Journal of Psychological Assessment* 25: 186–193.
- Supple AJ, Su J, Plunkett SW, et al. (2013) Factor structure of the Rosenberg Self-Esteem Scale. *Journal of Cross-Cultural Psychology* 4: 748–764.
- Van der Linden D, Te Nijenhuis J and Bakker AB (2010) The general factor of personality: A meta-analysis of Big Five inter-correlations and a criterion-related validity study. *Journal of Research in Personality* 44: 315–327.
- Venn JI (1980) On the diagrammatic and mechanical representation of propositions and reasonings. *Philosophical Magazine* 10: 1–18.
- Wright S (1921) Correlation and causation. *Journal of Agricultural Research* 20: 557–585.

Author biographies

Michael Dantlgraber studied Psychology at the University of Konstanz and worked there for several years as lecturer and researcher. He is an expert in statistics, psychometrics and test development.

Stefan Stieger, PhD, is a full professor for psychological research methods at the Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria. His work focuses on methods of Internet-based research, Ambulatory Assessment with smartphones and wearables, and open science.

Ulf-Dietrich Reips is a full professor at University of Konstanz. He works on Internet-based research methodologies, experimental design, measurement, personality, Social Media, causal knowledge. Ulf and his team develop and provide Web applications and methodologies for researchers, teachers, students, and the public (available from the iScience Server: <http://iscience.eu/>).