Introducing Item Pool Visualization (IPV)

Michael Dantlgraber[1], Stefan Stieger[2], & Ulf-Dietrich Reips[1]

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

2 …

Corresponding author

Correspondence concerning this article should be addressed to Michael Dantlgraber, University of Konstanz, Department of Psychology, Box 31, 78457 Konstanz, Germany

E-mail: michael.dantlgraber@uni-konstanz.de

Abstract

Keywords: data visualization, structural equation model, factor analysis, bifactor model, radex model, multitrait multimethod analysis

Introduction

In applied settings psychologists often face the challenge of selecting the right questionnaires suited to their specific questions. For example, if they want to use a questionnaire to measure individual differences, they have to choose between several well developed scales 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. In other words published scales differ regarding their content validity and consequently regarding a specific research questions they are more or less valid.

In this article we introduce Item Pool Visualization (IPV) and show how it facilitates scale comparisons. We visualize, for example, that two questionnaires that are both aiming to measure *self-esteem* have indeed a different understanding of this concept. While one questionnaire primarily may evaluate the level of self-esteem regarding the trust in social and task-related abilities, the other may primarily consider the lack of self-doubts (the lack of negative self-esteem).

IPV is based on comparisons of factor loadings when different structural equation models (SEMs; Wright 1921; Kaplan, 2001) are estimated with the same data. IPV is an illustration system that locates items and item pools (scales) regarding their similarities along several dimensions of nested radar charts. Item pools are visualized as circles that do not overlap. For this reason IPV substantially differs from traditional Venn diagrams (Venn, 1980) which also illustrate relations of variables with circles. In the next sections we systematically explain the structure of IPV and its advantages.

Methods and Results

Materials

In order to demonstrate the possibilities of IPV we used a large dataset containing three different 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, Golby, & van Wersch, 2009).

Sample

The sample was drawn in Germany and Austria and consists of 2,272 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. 1,265 (56%) participants 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 (*SD* = 17.7). Regarding highest educational level, 547 (24%) had a university degree, 783 (34%) had matura (high school graduation), 512 (23%) had an apprenticeship certificate, 333 (15%) had a mandatory degree, and 86 (4%) had no degree.

Procedure and analysis

IPV illustrates a comparison of different SEMs using the same data. 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 more specific item pools (e.g. facets). We show this by using the example of the DSSEI.

First, a general factor model is estimated. There, a single factor is extracted from an overall item pool including all items of the DSSEI (see Figure 1). Second, a correlated factor model is estimated.

There, four correlated factors are extracted from four smaller items pools representing the four facets of the DSSEI (*Social Competence*, *Task-Related Abilities*, *Public Presentation*, *Physical Appeal*; see Figure 2)

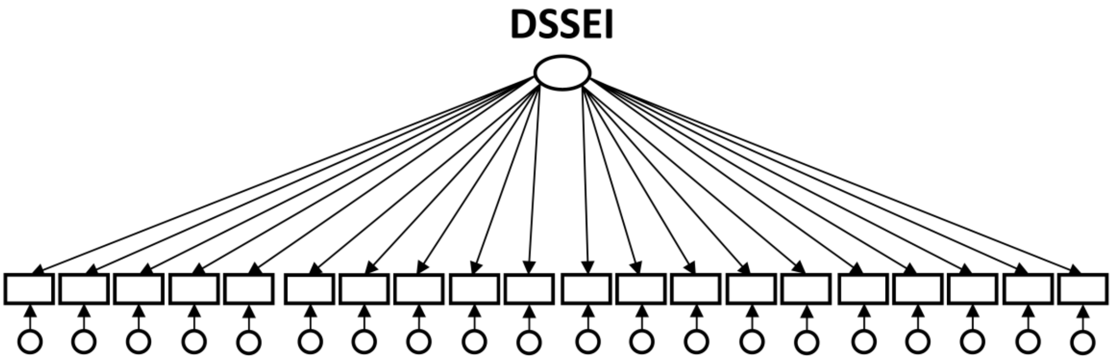

Figure 1. General factor model of the Domain Specific Self-Esteem Inventory (DSSEI).

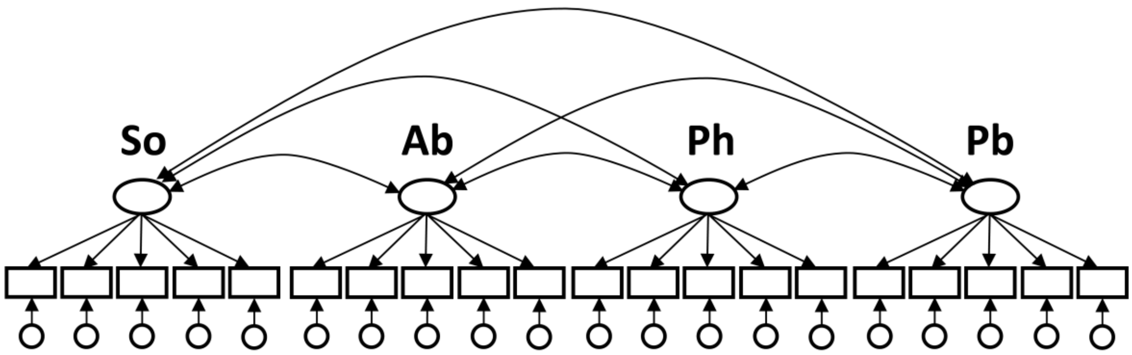

Figure 2. Correlated factor model of the Domain Specific Self-Esteem Inventory (DSSEI). So = Social Competence, Ab = Task-Related Abilities, Ph = Physical Appeal, Pb = Public Presentation.

When estimating the two models (Figure 1 and Figure 2), it is apparent that different factor loadings are estimated for the same items (see Table 1). This is because the factors are extracted from different item pools. Table 1 shows that the factor loadings of the correlated factor model tend to be higher than the loadings of the general factor model. This is due to the correlated factor model representing the facets, i.e. smaller and in most cases more specific item pools.

Table 1. Basic Calculation of the IPV of the DSSEI

Facet	Item	Factor loadings			
		Gen. factor model	Corr. factor model	Ratio of squared loadings	Center distance
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	0
	At social gatherings I am often withdrawn. not at all outgoing. (R)	.45	.43	0.91	0
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

Note. IPV = Item Pool Visualization, DSSEI = Domain Specific Self-Esteem Inventory, So = Social Competence, Ab = Task-Related Abilities, Ph = Physical Appeal, Pb = Public Presentation. Items followed by (R) were reverse-scored before analysis.

Visualizing items

In order to quantify the increase of the loadings when switching between the two models 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 one 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 item pool does not explain more variance of the respective item than the factor extracted from the overall item pool. In other words, a ratio of squared loadings equal to one represents no difference. 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 when the item is allocated to a smaller and more specific item pool instead of a larger reference pool. A center distance of 0 represents no increase. By definition, no increase includes a possible decrease of explained variance (i.e. negative values are treated as 0, see Table 1). For this reason center distances are never negative. Center distances are used for locating the items along several dimensions in a radar chart (see Figure 3). IPVs in this article are drawn by an R package by Petras and colleagues that is in development.

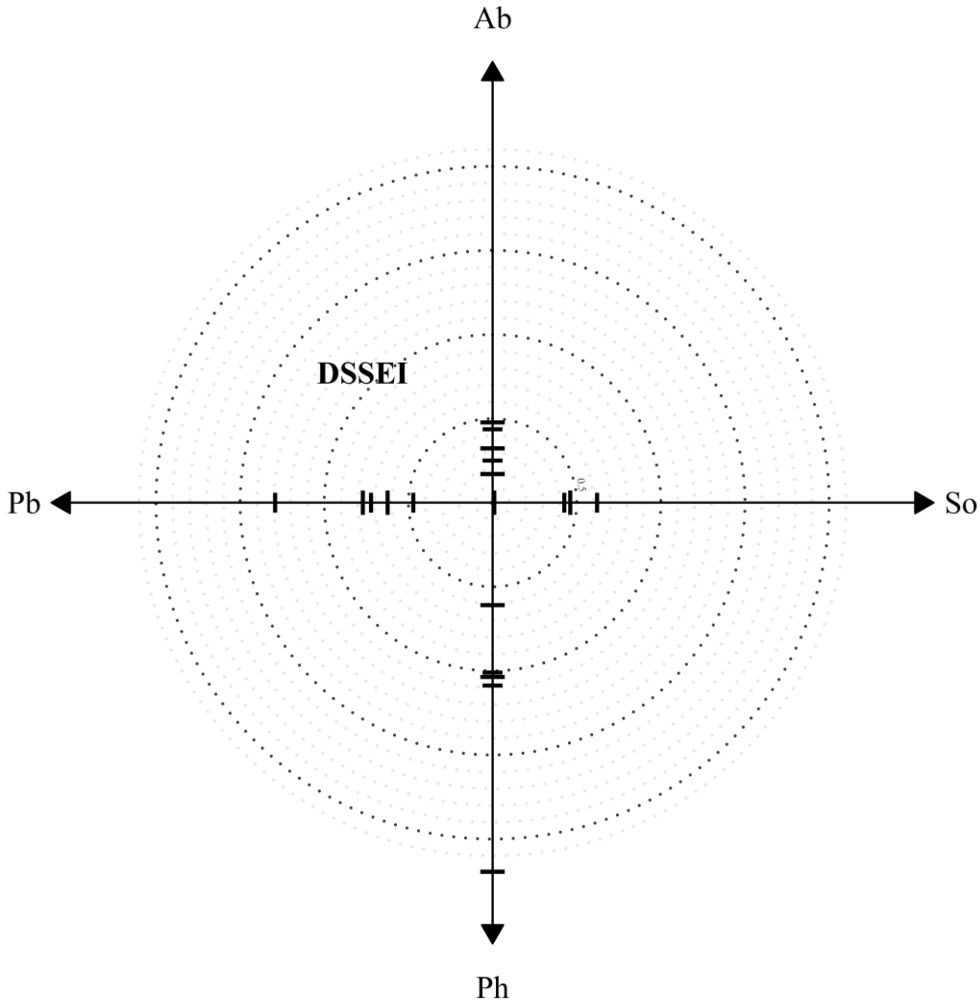

Figure 3. Item location of the DSSEI when using IPV. So = Social Competence, Ab = Task-Related Abilities, Ph = Physical Appeal, Pb = Public Presentation.

In an IPV the center of the radar chart represents the item variance that is explained by the factor extracted from the overall item pool. Therefore, center distances illustrate the proportional association of each item to a smaller item pool in comparison to 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

Items that are located on the same dimension can be combined to an item pool representing the respective facet. In Figure 4 the items are combined to four item pools representing the four facets of the DSSEI (*Social Competence*, *Task-Related Abilities*, *Public Presentation*, *Physical Appeal*).

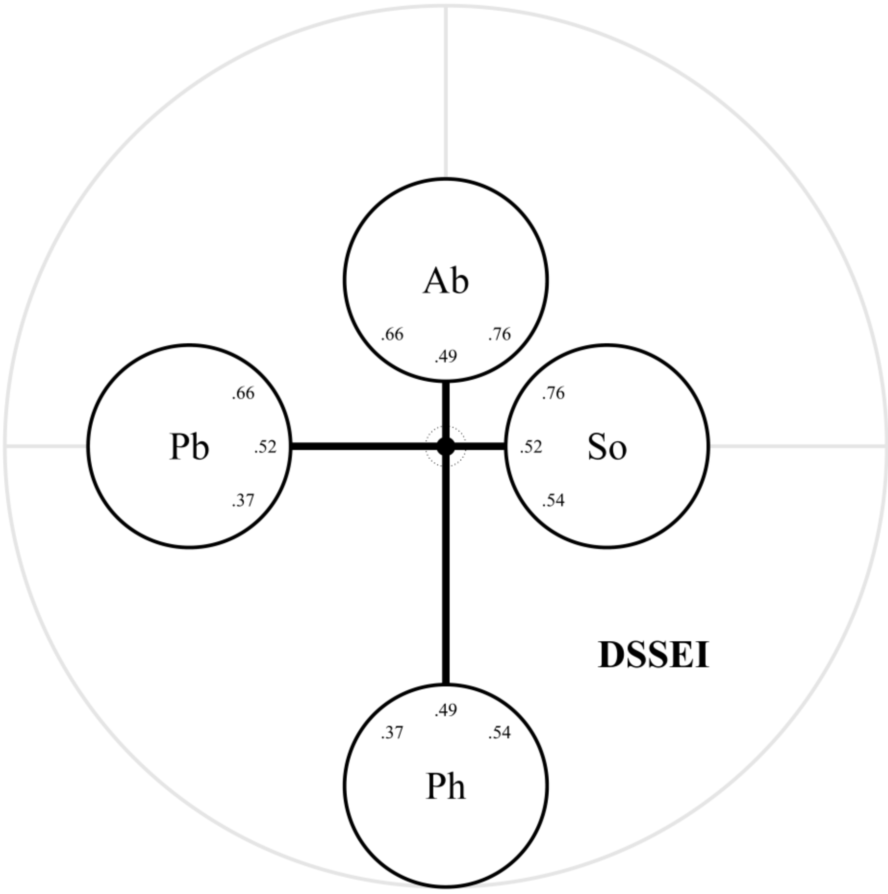

Figure 4. Facet location of the DSSEI when using IPV. So = Social Competence, Ab = Task-Related Abilities, Ph = Physical Appeal, Pb = Public Presentation.

The shifted circles representing the item pools 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. IPV uses this illustration to avoid confusing overlapping circles. Because the size of the circles does not have any meaning, it is always possible to adapt the scaling of the specific dimensions that the circles do not overlap.

The latent correlations of the factors extracted from theses pools (estimated with the model shown in Figure 2) are also included in Figure 4. That means that the numbers within a respective circle represent the latent correlations between this item pool (facet) and the other item pools (facets). IPV of the DSSEI shows that the facets are not balanced (see Figure 4). Contents of centered facets tend to be more relevant in the overall questionnaire. This is due to the fact that the contents of centered facets tend to be better represented by the general factor model, .i.e. they are also covered by items of the other facets.

Figure 4 shows that the centered facets *Social Competence*, *Task-Related Abilities* are correlated the most with the 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.

Requirements for IPV

IPV has three requirements. (1) The data must allow SEM. (2) All estimated factor loadings in each SEM should be positive. However, if all loadings of the same pool are negative, it is possible to recode the items and rename the pool (see *visualizing comparison of questionnaires*). (3) All estimated factor loadings in each SEM should not be too low. We recommend to exclude items that have factor loadings below 0.1, 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 comparison of questionnaires

As demonstrated in the previous sections IPV is a nested system that compares factor loadings when an item pool is split into some smaller pools. This system allows item pool comparisons on several hierarchical levels. Split item pools (facets) can be further split or overall item pools (questionnaires) can be combined with similar item pools (questionnaires) into a superordinate item pool. The latter is needed to compare the DSSEI with the RSES and the SMTQ that are also indicated to measure a positive self-concept. Such a comparison implies three hierarchical levels. Therefore three SEMs are estimated (see Figures 5-7).

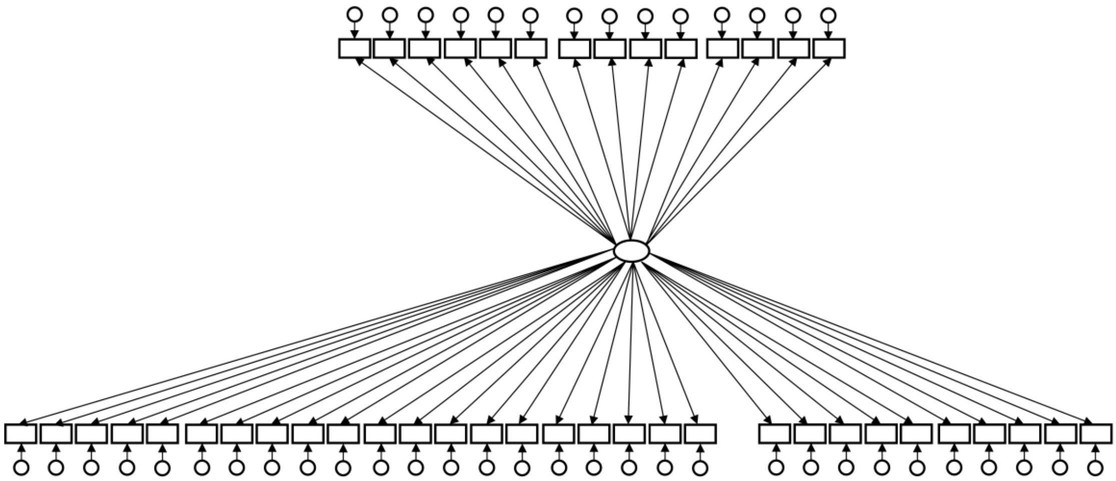

Figure 5. General factor model including all items of the DSSEI, SMTQ, and RSES.

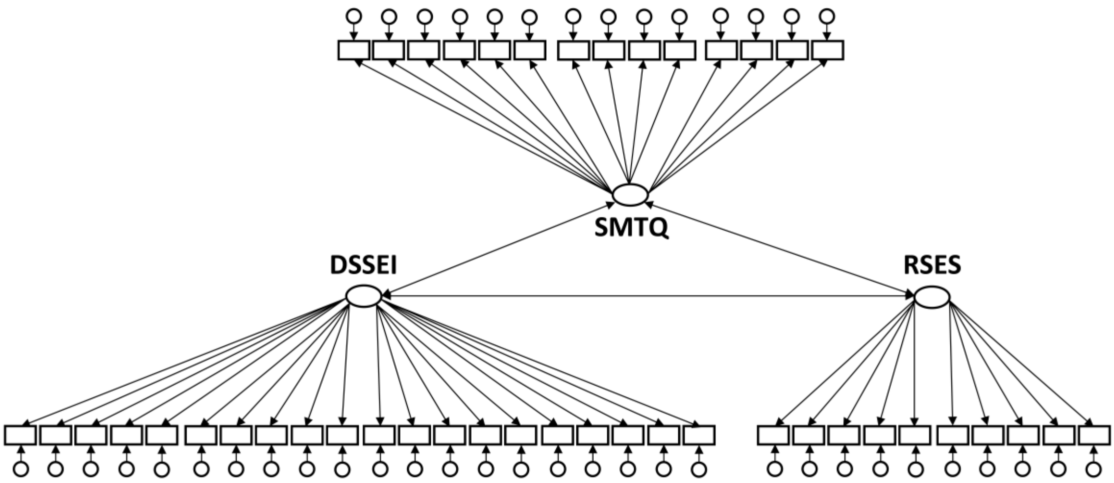

Figure 6. Correlated factor model of the main factors of the DSSEI, SMTQ, and RSES.

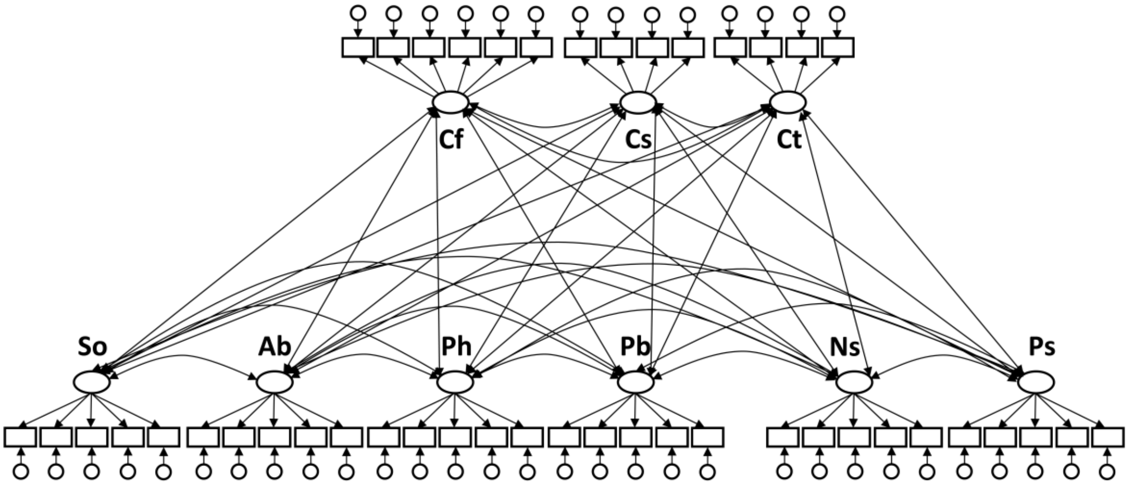

Figure 7. Correlated factor model of the facets of the DSSEI, SMTQ, and RSES. So = Social Competence, Ab = Task-Related Abilities, Ph = Physical Appeal, Pb = Public Presentation, Cf = Confidence, Cs = Constancy, Ct = Control, Ns = Lack of Negative Self-Esteem, Ps = Positive Self-Esteem.

In the first SEM (Figure 5) one factor is extracted from a superordinate item pool (from all items of the three questionnaires). In the second SEM (Figure 6) three correlated factors are extracted from three item pools representing the three questionnaires (DSSEI, RSES, SMTQ). In the third SEM (Figure 7) nine correlated factors are extracted from nine item pools representing the facets of the three questionnaires. Within the DSSEI and the SMTQ, we grouped items to facets as indicated by the original authors of the scales. 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, Su, Plunkett, Peterson, & Bush, 2013). In order to meet the requirements of an IPV (see *requirements of IPV section*), all items of the facet *negative self-esteem* were recoded and the facet was renamed to *Lack of Negative Self-Esteem.*

Due to the three model estimations (Figure 5-7) 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 (Figure 5 and Figure 6). Additionally, one center distance can be calculated that represents the comparison of the second and the third SEM (Figure 6 and Figure 7). When using *mean center distances* a complex IPV can be drawn that includes information of the three SEMs (see Figure 8).

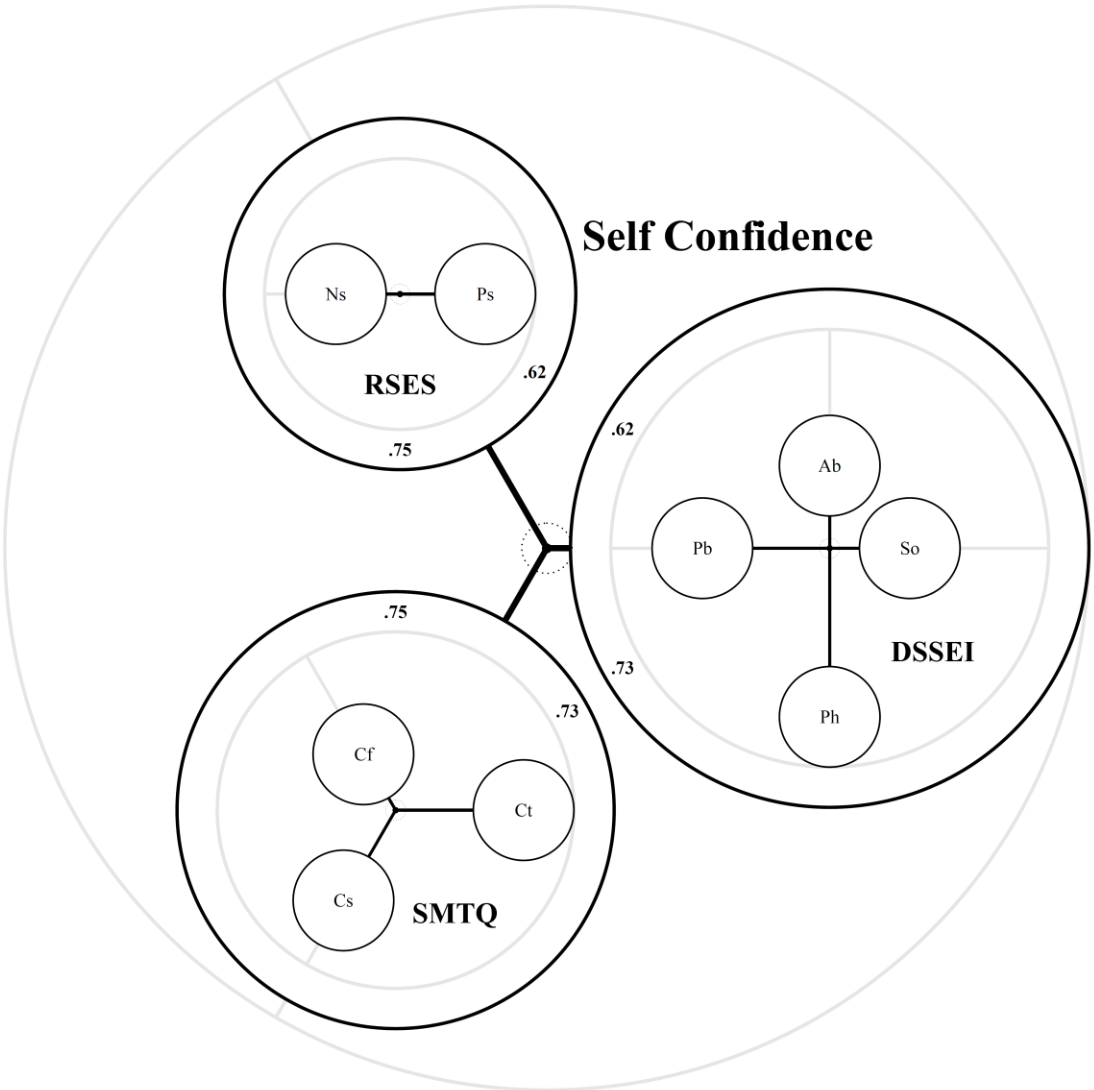

Figure 8. Complex IPV of the DSSEI, SMTQ, and RSES. So = Social Competence, Ab = Task-Related Abilities, Ph = Physical Appeal, Pb = Public Presentation, Cf = Confidence, Cs = Constancy, Ct = Control, Ns = Lack of Negative Self-Esteem, Ps = Positive Self-Esteem.

Figure 8 shows the comparison 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 6).

Note that the center distances within the three questionnaires are drawn on the same scale, but the center distances between the three questionnaires are twice that in order to avoid overlapping circles. The item “*In general, I feel confident about my abilities”* of the DSSEI facet *Task-Related Ability* and the item “*I feel confident about my social behavior”* of the DSSEI facet *Social Competence* have the highest loadings on the superordinate general factor, including all items of the three questionnaires (r = .66, r = .62). For this reason we named the superordinate factor (Figure 5) *Self-Confidence*.

Figure 8 integrates three radar charts 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 8). 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 fewest 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 fewest from the general factor of the DSSEI, which, as described above, differentiates fewest from the superordinate factor (*Self-Confidence*). This implies that *Social Competence* and *Task-Related Abilities* are strongly associated with *Self-Confidence* (the main source of variance of all investigated items).

In some cases, facets between different questionnaires are correlated more strongly than their general factors. 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 8 shows 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 factors

Using IPV facilitates the interpretation of factors. In Figure 8 the general factor of the RSES (representing *Self-Esteem*) shows a higher correlation to the general factor of the SMTQ (representing *Mental Toughness*) than to 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, i.e. are indicated to measure *Self-Esteem*.

However, Figure 8 also shows that the facet *Lack of Negative Self-Esteem* of the RSES differentiates lowest from its general factor (*Lack of Negative Self-Esteem*) is more centered than *Positive Self-Esteem*). Furthermore, the facet *Lack of Negative Self-Esteem* of the RSES shows a high correlation to the facet *Control* of the SMTQ. These results can explain the unexpectedly high correlation of their general factors. Interpreting Figure 8 facilitates the understanding of the questionnaires and their operational definition, i.e. IPV enhances the understanding what questionnaires really measure.

Interpretation rules for IPV

1. In an IPV circles represent item pools.
2. The only distances that have a meaning in a IPV are *center distances*, i.e. distances from the center of a certain radar chart to edges of other circles.

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 represents the mean proportional increase of the explained variance of the included items if they are allocated to a smaller item pool instead of a larger reference pool.
5. It is possible to define a route 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.
7. Dotted arrows represent the latent correlation between subordinate circles that do not share the identical superordinate circles, i.e. 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 rather the balancing of the facets. In most cases facets are not balanced what should be taken into account when interpreting the results of a questionnaire. 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 *Lack of Negative Self-Esteem* and therefore offers an opportunity for content related comparisons.

Furthermore, consisting of a nested system of item pool labels, IPV offers the opportunity to systematically proof existent labels. This is an important issue, because authors rarely name their scales on the basis of strict empirical rules (for example following the content of the items showing the highest factor loadings). Normally authors name their scales after an existent construct or theory whether their item pools are really representing it. For example, a scale that exclusively consists of items related to conspicuous party behavior is normally 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*.

In this regard it is important to note that this example describes a restriction of content validity that cannot be detected with measures of reliability or any kind of fit indices. Naturally, a *Conspicuous Party Behavior Scale* can be very reliable and can show very good fit indices, for example 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 what cannot be done by comparing measures of reliability or fit indices.

Relation to higher order factor models

IPV is based on comparisons of different SEMs that represent different hierarchical levels (see *procedure and analysis section*). 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 8. IPV has some advantages wherefore we prefer it.

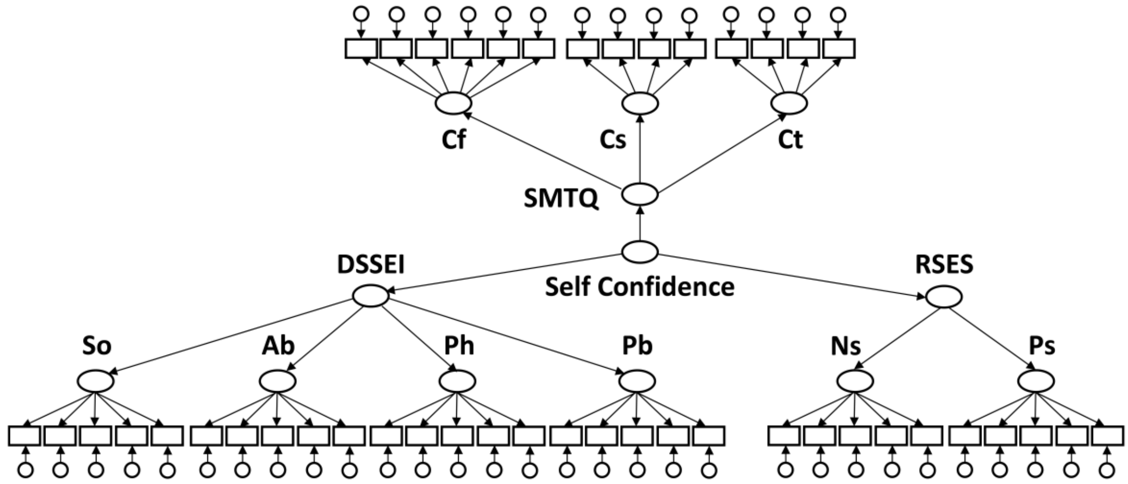

Figure 9. Higher order factor model of the DSSEI, SMTQ, and RSES. So = Social Competence, Ab = Task-Related Abilities, Ph = Physical Appeal, Pb = Public Presentation, Cf = Confidence, Cs = Constancy, Ct = Control, Ns = Lack of Negative Self-Esteem, Ps = Positive Self-Esteem.

- In a higher order factor model the arrangement of the questionnaires and facets have 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 pool.
- 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 7), which is, however, not hierarchical or nested. IPV (see Figure 8) combines the main 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 *requirements for IPV section*), IPV cannot be used to illustrate a complex network of 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 questionnaires, 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 items (this issue concerns all types of modelling).

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. IPV combines the main advantages of higher order and correlated factor models and offers the opportunity to locate items and questionnaires in large networks and to zoom in and out of these networks. IPV is well suited for proofing labels of questionnaires and facets and for comparing facets of different questionnaires.

References