Personality and Individual Differences 120 (2018) 288–298

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Facet Benchmarking: Advanced application of a new instrument refinement method

Alex B. Siegling *, Adrian Furnham, K.V. Petrides

Division of Psychology and Language Sciences, University College London, London, UK

ARTICLE INFO

Article history:
Received 27 August 2016
Received in revised form 2 November 2016
Accepted 9 December 2016
Available online 14 December 2016

Keywords:
Facet benchmarking
Scale construction and revision
Psychological measurement
Psychometrics
Dispositional mindfulness

ABSTRACT

This article presents an advanced application of Facet Benchmarking (FB), an instrument refinement method that sets out to identify *redundant* and *extraneous* facets (Siegling, Petrides, & Martskvishvili, 2015). FB uses external benchmarks to determine whether a measure's facets each occupy unique construct variance. In Study 1, three samples completed measures of dispositional mindfulness and an objectively derived set of construct-relevant criteria. A general factor extracted from these criteria was used to benchmark the measures' facets or subscales. Structural Equation Modelling, featuring a common latent (method) factor, was incorporated as an alternative statistical procedure, indicating that statistical or methodological artefacts were unlikely to account for the obtained results. Study 2 was conducted to cross-validate the results for a benchmark derived from a different set of criteria. The results support the method's robustness and efficacy.

1. Introduction

A core challenge with psychological constructs is to accurately determine their domain of measurable manifestations, or construct domain. This process is often facilitated by the explication of facets, especially where broader constructs are concerned. Determining construct domains involves considerable uncertainty (Costa & McCrae, 1998; Ziegler, Booth, & Bensch, 2013), since an individual and objective criterion against which measures can be evaluated does not normally exist (Epstein, 1984; John & Benet-Martinez, 2000). Various psychometric paradigms (e.g., construct validity vs. test validity theory; Borsboom, Mellenbergh, & van Heerden, 2004) and statistical procedures (e.g., Exploratory Structural Equation Modelling, Bifactor Modelling) have enriched psychometrics, but the process of operationalising constructs remains far from clear-cut (Ziegler & Bäckström, 2016). Consequently, one encounters a diversification of measures as well as an overall plethora of facets for many constructs (Pace & Brannick, 2010).

Relevant substantive approaches specifically concerned with the explication of facets and testing multi-faceted constructs have emerged within recent decades (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Costa & McCrae, 1998; Hull, Lehn, & Tedlie, 1991). To various extents, available item-selection and -evaluation procedures can also be applied to the assessment of facets (see Smith, Fischer, & Fister, 2003). The problem is that the available approaches were not developed with the aim of identifying problem facets detrimental to validity, viz. redundant facets and, to a lesser extent, extraneous facets (see Siegling, Petrides, & Martskvishvili, 2015, for a more detailed conceptualisation of problem facets). The decisive characteristic of redundant and extraneous facets is that neither of them represent unique elements of the target construct; extraneous facets represent no elements whatsoever. It is this characteristic that (the authors contend) the conventional validation and scale development approaches cannot meticulously unveil.

Although much progress has been made to disentangle different sources of variance statistically (Morin et al., 2016; Raykov & Marcoulides, 2016; Schmid & Leiman, 1957), the reliable identification of redundant and extraneous facets, based on their inability to occupy unique construct variance, is not simply a matter of statistics. It depends heavily on the selection of variables and input data. Facets are typically evaluated against one another (i.e., along with variables characterised by a similar level of uncertainty). If a set of facets represents the target construct poorly, extraneous facets are more likely to load on the latent variable, and examining multicollinearity (e.g., in confirmatory factor analysis) is no trustworthy approach to detecting redundant facets. It is, thus, risky to assume that even advanced statistical procedures reliably distinguish the real target construct from other constructs as well as redundant construct variance from unique construct variance. Importantly, the concept of unique construct variance differs from specific variance; the former refers to a facet's unique part of the target construct and the latter to the part that is unrelated to the construct (see Fig. 1).

This article further examines the efficacy of Facet Benchmarking (FB), a recently proposed instrument refinement method that sets out to identify redundant and extraneous facets systematically (Siegling et

* Corresponding author at: London Psychometric Laboratory, University College London, WC1H 0AP, UK.
E-mail address: a.siegling@ucl.ac.uk (A.B. Siegling).

http://dx.doi.org/10.1016/j.paid.2016.12.014
0191-8869/

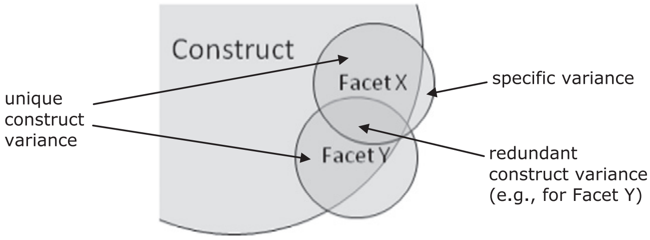

Fig. 1. Decomposition of construct variance into unique and redundant construct variance.

al., 2015). A concise description and advanced application guidelines are given next.

1.1. Facet Benchmarking (FB)

The concept of criterion validity has relevance in the identification of redundant and extraneous facets. Without unique construct variance, a facet is less likely to explain (unique) variance in construct-relevant criteria. Also, correlations of a scale composite that encompasses problem facets with construct-relevant criteria are systematically, although not necessarily always, lower than those of a composite comprised only of valid facets (see Smith et al., 2003, for a more detailed discussion of this effect). The construct-unrelated variance imposed on the composite by extraneous facets further compromises the composite's criterion validity (where construct-relevant criteria are concerned). Smith et al. have discussed how the external approach, or criterion keying, can be extended by means of incremental validity principles, with the aim of identifying and retaining facets with unique explanatory effects. However, the pivotal question not addressed in their seminal article concerns the criteria to be used for assessing the incremental value of individual facets, or whether each facet occupies unique construct variance.

One issue in leveraging criteria for the purpose of assessing facets is that, individually, they are unlikely to qualify as a comprehensive construct representation (Epstein, 1984; John & Benet-Martinez, 2000). Furthermore, like facets, individual criteria can comprise specific variance unrelated to the target construct; they are often multidimensional and cannot be expected to represent the construct variance exclusively (Smith & Zapolski, 2009). Due to sources of variance other than the target construct, there would be an increased chance of seeing explanatory effects of extraneous facets and, to a lesser extent, redundant facets. It also is realistic that some facets correlate positively with a particular criterion, whilst other facets correlate negatively with the same criterion (Ziegler, Danay, Schoelmreich, & Buehner, 2010).

As a remedy to the difficulties, individual benchmarks can be objectively derived from the shared variance of representative and balanced sets of construct-relevant criteria, selected with the construct as a reference point. Precisely, such a latent variable may be viewed as an approximation of the construct variance, with its accuracy depending on the method of derivation and knowledge about the construct already existing. In a five-stage process, FB examines whether a facet can occupy a unique portion of variance in these benchmarks.

1.1.1. Stage 1

The challenge is to select a set of construct-relevant criteria that represents the construct variance comprehensively (i.e., not missing any variance) and exclusively (i.e., not imposing variance unrelated to the construct). While both these requirements inevitably involve a theoretical process, exclusiveness is considerably facilitated by the statistical procedures described at Stage 2. Comprehensiveness is facilitated by incorporating varying, systematically selected sets of criteria, if necessary. Ideally, one would obtain a representative sample of all construct-relevant criteria without duplicating any elements, thus aiming for a balanced representation (it seems undesirable to use all conceivable criteria, since many of them are likely to overlap in their construct-related variance). If the benchmark is unbalanced with respect to the construct, the construct variance represented would shift towards individual facets, which can bias the FB results.[1]

Perhaps most straightforward is to rely on variables conceptualised as at least partial, direct psychological outcomes and, perhaps, known to correlate in the expected direction with the target construct. Indirectly-related outcomes increase the chances of significant explanatory effects of extraneous facets, since these are less likely to represent the target construct primarily. Although, prior empirical correlations may not be necessary, and other, more theory-driven approaches may be incorporated in making these decisions. Another consideration warranted during the criteria selection process are situational moderators, which can influence facet-criterion relationships. For instance, the central tenet of Trait Activation Theory is that situational factors (e.g., job demands, distractors) influence the expression of personality traits and their associations with relevant outcomes (Tett & Burnett, 2003). It is vital that the chosen criteria are either relevant across situations (i.e., general) or systematically sampled from all pertinent situations.

1.1.2. Stages 2 and 3

The basic idea is to extract the first latent factor, or benchmark, from the criteria administered to each sample, and then examine which facets occupy unique variance within this benchmark. There are two salient options for execution: (1) separately via factor analysis[2] and multiple regression; and (2) jointly via Structural Equation Modelling (SEM). Partially, the procedural choice depends on context considerations, such as sample size and number of facets and criteria involved.

1.1.2.1. The fragmented procedure: factor analysis and regression. The first latent factor is, in theory, the variable representing the target construct, since the criteria were selected using the construct as the reference point. As a proxy representation of the homogenous construct, the benchmark (an alternative latent variable of the construct variance) is most appropriately extracted via principal axis factoring (1 factor, no rotation), although we have previously used Principal Component Analysis for this purpose (the two procedures tend to yield virtually identical results for the first variable extracted). Any unrelated criteria (i.e., those that do not load well on the first factor) are identified and excluded in this process.

The question is at what threshold to drop or retain a criterion. Any divergent criteria may still co-vary due to sources other than the target construct, such as common method effects or chance. Consequently, they can introduce construct-unrelated variance on the benchmark. On the other hand, there is a danger of dropping valid criteria of the construct. For the time being, it makes sense to proceed with a generic minimum loading of 0.30, the common cut-off for scale items or facets in scale construction. A pre-specified value is intended to foster reliability and replicability of results, although it may be unwise to strictly advocate a specific cut-off. The important point is that adjustments are made a priori, guided by reason and theory.

Stage 3 of FB examines whether each of the facets occupies unique variance within the derived benchmark and if the variance explained is in the expected direction. A suitable statistical procedure for this purpose is statistical regression (also referred to as the stepwise method), with all facets entered at the initial step as explanatory variables of the benchmark. Stepwise regression both removes (criterion:

[1] Although a balanced representation of the construct is desirable in the context of FB, it is does not seem fatal if some criteria included within the benchmark are redundant with one another (if they do not share any specific or error variance with redundant or extraneous facets). Most redundant facets will be unable to account for unique construct variance, irrespective of whether their redundant variance is duplicated within an objectively derived benchmark.

[2] Note that the general limitations of factor analysis as a stand-alone statistical procedure in screening out redundant and extraneous facets are compensated within the context of FB.