FACET BENCHMARKING:
A PSYCHOMETRIC METHOD FOR REFINING MULTI-FACETED ASSESSMENT INSTRUMENTS

By

ALEX B. SIEGLING

Prepared under the supervision of:

Dr. Konstantinos V. "Dino" Petrides

and Prof. Adrian F. Furnham

A thesis submitted for the degree of Doctor of Philosophy

Division of Psychology and Language Sciences

University College London

July 2015

SIGNED DECLARATION

I, Alex B. Siegling, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

Driven by the challenge of defining and measuring psychological attributes, this thesis advances an instrument refinement method aimed at identifying "problem" facets detrimental to construct validity. The method, labelled "Facet Benchmarking" (FB), integrates theoretical and empirical steps and is intended to supplement established scale construction approaches; it is part of the wider construct validation paradigm. FB seeks to detect *redundant* and *extraneous* facets based on their inability to occupy a unique part of the variance attributed to a given construct. An alternative, more objectively derived representation of the construct is used to assess if the hypothetical facets of a given measure fulfil this general criterion. That representation is a composite extracted from systematically selected criteria, or outcomes, of the construct.

In this thesis, FB is examined across three investigations (three chapters) of increasing rigour, each involving a different construct and data from multiple samples. The first application of FB (**Chapter 2**) was based on existing data, gathered in previous validation studies of the Trait Emotional Intelligence Questionnaire. **Chapter 3** lays the psychometric groundwork for the second application of FB, addressing issues of homogeneity and dimensionality of the relevant measures. These were measures of dispositional mindfulness, the construct to which FB is applied in **Chapter 4**. The third application of FB focuses on the General Factor of Motivation, a re-conceptualisation of motivation proposed and validated in **Chapter 5** (two measures were developed and used for this purpose). The purpose of this final investigation (**Chapter 6)** was to assess plausible alternative explanations for the method's efficacy: domain underrepresentation and common-method variance between facets and criteria. The results from all three investigations of FB supported the efficacy and integrity of FB.

The implications are, therefore, discussed in detail in **Chapter 7**, along with considerations for the method's application and future development.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: General Introduction

Examining the substantive literature of a psychological construct, one often encounters a diversification of psychometric measures as well as an overall plethora of facets used to represent the construct. In some cases, the arrays of facets used to represent the same construct diverge considerably (in quantity and/or types) and correlations between measures are only weak or moderate (Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006; Brackett & Mayer, 2003; Matthias Ziegler, Booth, & Bensch, 2013). It is then difficult to accept that all measures tap into the same underlying attribute accurately, or evidence construct validity. The existence and use of multiple measures complicates the comparison and aggregation of research findings, particularly if these measures (and their underlying models) vary substantially. In applied contexts, where psychometric assessment constitutes a major pillar of comprehensive psychological assessments, measures of poor construct validity can invalidate assessment results, impinging negatively on the individual assessed and society at large.

The challenges associated with defining and measuring psychological constructs have been noted for many decades. Psychometrics, the science of scale construction behind psychological assessment, sets out to represent and measure psychological constructs indirectly, based on their cognitive, affective, and behavioural manifestations (Cronbach & Meehl, 1955; Raykov & Marcoulides, 2011). The focus on the "observable" is the only gateway to studying and measuring psychological constructs, and it is not clear when and if more straightforward methods will emerge (e.g., from Neuroscience). This epistemological constraint is the fundamental reason why the definition and measurement of psychological constructs is so challenging.

On the one hand, psychometric methods have shown invaluable utility in advancing psychological research and applications. Considerable progress has been achieved over the years, and various paradigms (classical test theory, generalisability theory, and construct validity theory), approaches (e.g., Cronbach & Meehl's, 1955, nomological network and Campbell & Fiske's, 1959, multitrait-multimethod matrix), and statistical innovations (e.g., factor analysis, structural equation modelling, and item response theory), have enriched, and contributed to the advancement of, psychological measurement. Still, the process of conceptualising and operationalising constructs remains far from straightforward, and one must ask if the available psychometric approaches and tools are optimal and sufficient. Like any other area of scientific enquiry, psychometrics should be viewed as a self-learning system that grows with theoretical, statistical, empirical, and technological advancements.

A core challenge with psychological constructs is how to define them and accurately represent their domain of manifestations (construct domain), or to sample all the relevant content. This process concerns Loevinger's (1957) notion of "substantive validity" and is often facilitated by the explication of facets, especially where broader constructs are concerned. Defining the construct domain (e.g., by specifying the facets) to represent a construct and ascertain accurate measurement involves considerable uncertainty (Costa & McCrae, 1998; Petrides & Furnham, 2001; Matthias Ziegler et al., 2013), since an individual and objective criterion against which measures can be evaluated does not exist for most constructs (Epstein, 1984; John & Benet-Martinez, 2000; John & Soto, 2007). Psychological constructs are inherently "fuzzy" and lack clear-cut boundaries (Matthias Ziegler, Kemper, & Lenzner, 2015). However, as explained later in this introduction, some of this difficulty stems from a lack of efficacy of the contemporary scale construction approaches in evaluating facets and, specifically, in screening out "problem facets" (i.e., facets that compromise the validity of a given measure or representation). Consequently, they are unable to prevent the inflation in the overall number of facets and diversification of measures seen in the literature of many (if not most) constructs.

This thesis describes and examines a new psychometric method for refining multi-faceted assessment instruments. Instrument refinement has been defined as "any set of procedures performed on an instrument designed to improve its representation of a construct" (Smith & McCarthy, 1995, p. 301). On the one hand, it heavily overlaps with scale construction, being applied during the original construction process, as post-hoc modifications, as well as for repurposing instruments for a new use. On the other hand, instrument refinement was historically neglected or lacked rigour and appropriate procedures (Smith & McCarthy, 1995). Although primarily considered an operationally focused endeavour, both instrument refinement and scale construction also have broader, substantive implications for the conceptualisation and definition of constructs (Cronbach & Meehl, 1955; Smith & McCarthy, 1995), as discussed in more detail in Section 1.2. In fact, scale construction (measurement) has been defined as "a process of building models that represent the phenomena of interest" (John & Benet-Martinez, 2000, p. 339).

Intended to supplement the contemporary approaches to scale construction and refinement, the proposed method targets problem facets detrimental to construct validity. It operates on the principle that problem facets are unable to occupy a unique part of the target construct's variance (to be delineated later in the present chapter), using an alternative representation of the construct as a benchmark to assess whether a measure's facets fulfil this general criterion. Given its focus on facets, the method has been termed "Facet Benchmarking" (FB). Since scale development and validation go hand-in-hand with our understanding of psychological attributes (theory building), FB also has direct implications for the development of construct representations and can help maximise the level of objectivity and minimise discrepancies in psychological measurement.

To avoid any conceptual uncertainty, this introductory chapter first presents definitions of key psychometric terminology used in this thesis that is often applied differently between authors. Second, the current paradigm in psychometrics, construct validation, is reviewed, featuring an overview of the existing approaches to scale construction. Third, problem facets are conceptualised with reference to essential guiding criteria a facet should satisfy (to qualify as a valid facet), and the limitations of the established psychometric approaches in identifying these facets are described. Fourth, the unique purpose of FB is described and its application is explained as a step-by-step procedure. The final section of this introduction outlines the empirical chapters of this thesis, which examine the efficacy and credibility of FB.

1.1. Definitions

The term *dimension* is used a technical synonym for "construct" or "domain"; it connotes that a variable (e.g., extraversion) is distinct from other, mostly non-overlapping variables, or dimensions (, e.g., neuroticism), rather than being the same construct, or a subfactor thereof. However, a single dimension may still be multi-faceted (as opposed to multidimensional). Especially in instances where the focal construct is relatively broad, the construct domain is frequently represented by a set of *facets,* which help ascertain that all relevant content areas are represented in a measure. Similar to previous definitions (Costa Jr. & McCrae, 1995), facets are viewed here as theoretically derived variables used to represent a single construct, which is reflected in the facets' common variance. Facets are interrelated variables that represent narrow and homogenous subsets of affective, behavioural, or cognitive manifestations (in psychometric terms items) of a given construct. Statistically, facets are often modelled as indicators of a latent construct or a first-order factor. The term *subscale* can be used to refer to any type of composite score of a given measure other than the global composite, including facet and factor scores. However, this term is reserved here for scales that are part of the same measure, but which do not yield a higher-order factor, in order to distinguish these variables from facets.

1.2. Psychometrics: Current Paradigm and Contemporary Approaches

Scale construction, including refinement, is seen as a process of developing theory, or measurement models, and vice versa; one cannot separate the two (John & Benet-Martinez, 2000; John & Soto, 2007; Simms & Watson, 2007). As Smith (2005) has stated succinctly: "to validate a measure of a construct is to validate a theory" (p. 413). This paradigm and process has become established as "construct validation". A central tenet of construct validation is that specific theories describing relations among psychological processes be specified and the performance of the focal measure against these theories be evaluated (Cronbach & Meehl, 1955; Strauss & Smith, 2009). Alternative conditions for test validity were proposed by Borsboom, Mellenbergh, and van Heerden (2004): (a) the attribute must exist, and (b) variations in the attribute must causally produce variation in the measurement outcomes.

Construct validation spans the validation of both measures and underlying theories, and it guides the development of new ones (Clark & Watson, 1995; Loevinger, 1957; Messick, 1995; Watson, 2012). Initially proposed by Cronbach and Meehl (1955), the paradigm has been subsequently advanced and elaborated by others (e.g., Loevinger, 1957; Messick, 1995), who have proposed different phases of construct validation, each seeking to address different aspects of the process. For instance, Loevinger emphasised construct validity as the superordinate concept (over specific types of validity and reliability) and proposed construct validation as the general framework for developing measures and theories. This framework is divided into three phases (substantive validity, structural validity, and external validity) and continues to guide scale construction in the present day (accessible practical guidance for applying Loevinger's framework is given in Clark & Watson, 1995). Of note, construct validation is considered a process that is ongoing and indeterminate, necessitating refinement and replication (Cronbach, 1988; Grimm & Widaman, 2012; John & Soto, 2007; Smith & McCarthy, 1995; Smith & Zapolski, 2009; Watson, 2012).

Construct validation is a circular rather than linear process. It begins with a theoretical process that focuses on the conceptualisation and definition of the construct, and on the development of an initial item pool (Loevinger's, 1957, notion of substantive validity). With implications for all other aspects of the construct validation process, clearly defining the target construct and embedding it within a nomological network is of utmost importance (Messick, 1995; Nunnally & Bernstein, 1994; Matthias Ziegler et al., 2013; Matthias Ziegler, 2014b). This step often involves the explication of facets, the focus of the present thesis. Once a satisfactory definition or representation has been identified, scale developers proceed to the selection of items for the measure (Loevinger's, 1957, structural validity phase), aiming for structural and discriminant validity. Finally, associations of the scale score with measures of the same, related, or entirely different constructs are examined to evaluate convergent, criterion, and discriminant validity, respectively (Loevinger's, 1957, external validity phase), ideally within a multitrait-multimethod matrix framework (Campbell & Fiske, 1959; Matthias Ziegler et al., 2013). The primary focus of this phase is on evaluation, rather than construction. Nonetheless, results from both of the two empirical (internal and external) phases can inform the validity and possible refinement of the operationalisation and representation of the construct.

Domain sampling and definition, which is of concern in the substantive validity phase, is mainly a theory-driven process, although often involving some form of qualitative research. In contrast, various theoretical, qualitative, and quantitative strategies for item selection (addressing structural validity) have emerged over the decades, though they can all be classified into one of three categories: the rational-intuitive, internal, and external approaches (Burisch, 1984; John & Benet-Martinez, 2000; Simms & Watson, 2007). Comparison studies of these approaches have generally found them to be equally effective (Burisch, 1984; Hase & Goldberg, 1967). Although excellent descriptions and guidelines for their application can be found elsewhere (e.g., John & Benet-Martinez, 2000; Simms & Watson, 2007), a general overview of these approaches is imperative here. Prior to describing each approach, it is important to stress that these categories of approaches are most wisely used in conjunction, given the unique strengths and limitations of each. In modern psychometrics, the rational-intuitive approach and the internal approach are nearly always involved, whereas the external approach is used infrequently.

1.2.1. Rational-intuitive approach

The focus of the rational-intuitive approach (also known as deductive approach) is strictly on theory and reasoning (very similar to the domain sampling procedures used in the substantive validity phase). Items are generated and scrutinised on the basis of theory, reflecting scale developers' theoretical understanding of the target construct. In conjunction with other approaches, the rational-intuitive approach maintains widespread appeal, featuring advanced qualitative methods. Examples include content analysis, having experts rate the relevance and representativeness of items (Buss & Craik, 1983; Haynes, Richard, & Kubany, 1995), and seeking consensus from trained raters in order to ascertain the construct fit of items and content homogeneity of facets (Harkness, McNulty, & Ben-Porath, 1995; Smith & McCarthy, 1995). Some form of theory in scale construction seems inevitable (Clark & Watson, 1995; Loevinger, 1957). Pertinent types of evidence the approach aims to establish are content and face validity. While measures based purely on this approach also tend to evidence decent convergent validity, evidence for discriminant validity tends to be weak (Simms & Watson, 2007). A general limitation of the approach is the unrealistic premise that the scale developer's understanding of the construct is correct.

1.2.2. Internal approach

The internal approach (also known as internal consistency or inductive approach) concerns the identification of homogenous scales (factors) and factor structure among items, thereby examining and establishing factorial and discriminant validity. Naturally, the approach also addresses internal consistency at different scale levels so as to remedy unnecessary error variance (Smith & McCarthy, 1995). It involves techniques grounded in classical test theory, factor analysis, and item response theory (Simms & Watson, 2007). The current trend and best strategy in evaluating and improving structural models is comparative model testing (John & Benet-Martinez, 2000; John & Soto, 2007), which compares the model fit of theoretically plausible alternatives. These models can differ in various respects, such as in the number of factors or factorial structure (e.g., hierarchical versus correlated factors). Given its pure empirical basis, the approach is unable to label the extracted factors, which illustrates the need for theory and integration with other methods.

1.2.3. External approach

This approach (also known as criterion-keying) focuses on criterion validity, by selecting items on the basis of their ability to discriminate between a relevant criterion group and a normal (control) group (i.e., groups differing on the target attribute). For example, in a clinical context, items may be selected if they are able to discriminate between groups of individuals differentiated by a clinical diagnosis, based on diagnostic criteria. The discriminating items are then included in the scale. In terms of construct validation, an attractive outcome of this approach is that criterion validity is automatically built into the resulting measure. However, with item content being virtually irrelevant, the approach has been criticised for its lack of theory in selecting items (e.g., Loevinger, 1957), several of which may be conceptually unrelated to the construct. It also is prone to producing heterogeneous scales, complicating the interpretation of scores and impinging negatively on discriminant validity (Smith, Fischer, & Fister, 2003). Yet, the selected items tend to exhibit large correlations, indicating redundancy. Even though the sole use of the external approach can hardly be defended, its utility is still recognised, particularly if used in combination with other item-selection strategies.

1.3. Problem facets

1.3.1. Conceptualisation

As a first step towards conceptualising problem facets (and distinguishing them from valid facets), it is imperative to define the criteria a variable should satisfy to qualify as a useful facet of a higher-order construct. Three basic criteria that a valid facet should satisfy can be specified. Fulfilment of these criteria would render a variable useful as a facet of a higher-order construct. On the other hand, failing to satisfy any of these criteria would cast doubt on its validity as a facet.

> Criterion A: A facet must tap into a homogenous set of psychological processes, situated at the same level of abstraction as the other facets of the measure. Specifically, a facet should represent a most specific set of psychological manifestations, or attribute, of the construct, rather than a distant outcome indirectly influenced by it (e.g., number of friends or romantic partners, highest level of education achieved, or age of death), a higher-order factor between the latent construct and facets (e.g., one of the 10 aspects of the Big Five, rather than a facet; DeYoung, Quilty, & Peterson, 2007), or even an antecedent causal variable (e.g., parenting style). This criterion is primarily addressed by theory and basic, non-psychometric research.

> Criterion B: A facet should share a non-negligible amount of variance with the other facets (i.e., common variance). The reason is that unwanted sources of variance, such as other dimensions or method and response biases (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), could account for modest correlations between facets and facet loadings on the latent composite. This criterion is the focus of factor analysis. However, although often taken as such, the possession of common variance is insufficient as the sole empirical criterion for the validity of facets.

Criterion C: A facet should occupy a unique portion of the variance attributed to the construct of interest (i.e., common variance not covered by other facets within the construct representation). This criterion remains unaddressed by the existing psychometric approaches and is the main focus of FB.

As regards to Criteria B and C, two types of problem facets can be operationally defined and referred to as *extraneous* and *redundant* facets. The best way to describe these facets is with respect to their component variance, as graphically illustrated in Figure 1.1 and Figure *1.2*. Facets can have two types of systematic variance: reliable common variance, which is due to the target construct and shared with the other facets, and reliable specific variance, which is unrelated to the target construct (e.g., Smith et al., 2003). In regards to the target construct, extraneous facets lack common variance; their variance is due to sources other than the target construct, a violation of Criterion B. Yet, extraneous facets may still share variance with valid facets, because of measurement bias or specific variance (shared dimensions other than the target construct). Redundant facets have common, construct-related variance, but this variance is already covered more efficiently by one or more of the other facets within the construct representation. In other words, redundant facets lack "unique common variance" and, therefore, do not add to the representation of the construct (Criterion C). In short, neither redundant nor extraneous facets represent a unique part of the construct variance, given all other hypothetical facets.

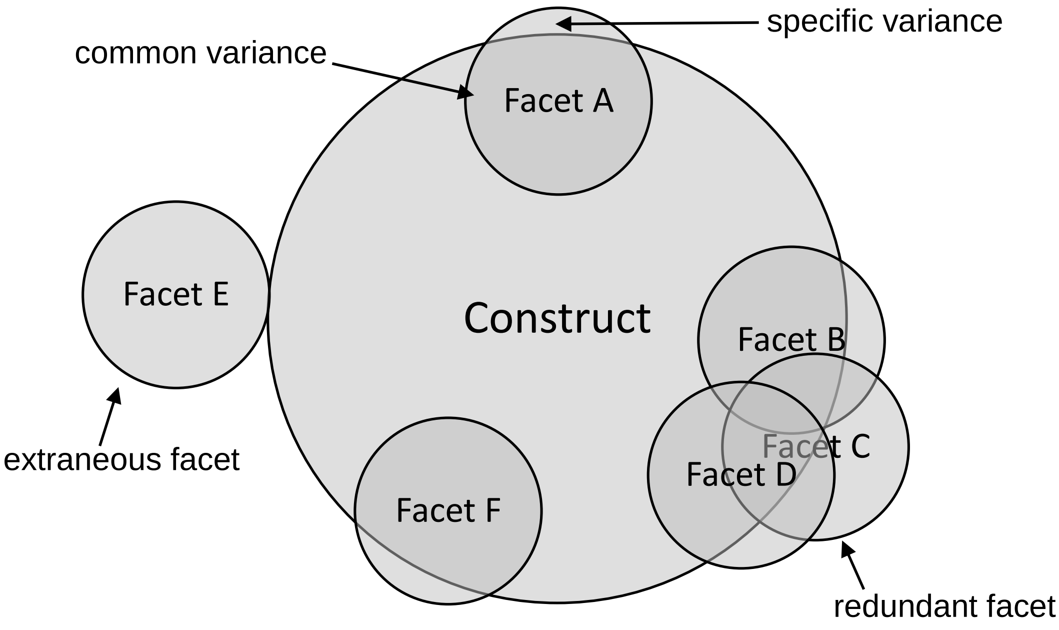


Figure 1.1. Illustration of redundant and extraneous facets with respect to their component (i.e., common and specific) variance.

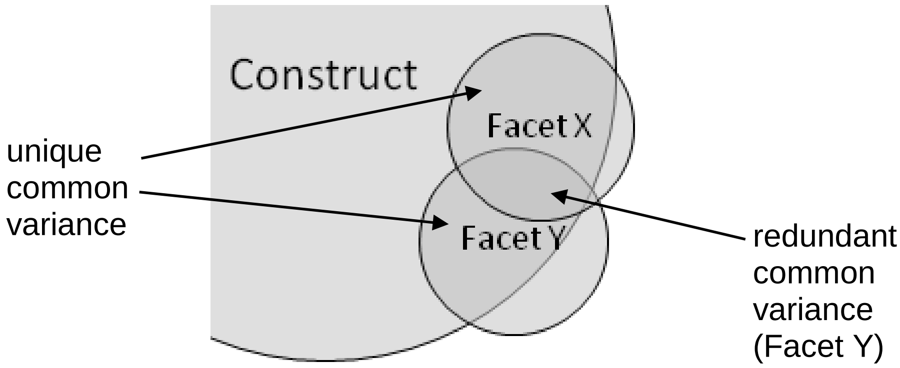


Figure 1.2. Decomposition of common variance into unique and redundant common variance.

Rather than contributing to the representation of the construct of interest, both these types of facet compromise the construct validity of a model or measure, which will be reflected in the various empirically testable subtypes of validity (concurrent, predictive, discriminant, etc.). Redundant facets lead to some manifestations of the construct being overrepresented and, consequently, to an overall unbalanced representation of the construct variance. Extraneous facets result in representations that exceed the target construct's boundaries, representing manifestations of other, non-targeted dimensions.

The empirical effect of these facets is that they both compromise the validity of the global composite systematically. Neither is uniquely representative of the target construct and, hence, unlikely to occupy a distinctive portion of its variance vis-à-vis the other facets. Consequently, the correlations of the total scale composite with proximate, construct-relevant criteria are systematically, although not necessarily always, lower than those of a composite without these facets (i.e., a composite comprised exclusively of facets that occupy a unique portion of the construct variance, which redundant and extraneous facets are lacking). The reason is that, without unique construct variance, a facet is unlikely to predict unique variance in construct-relevant criteria, with the effects of predictive and non-predictive facets averaging out when combined into a global composite (Smith et al., 2003). Moreover, since extraneous facets stretch the variance of the composite thought to represent the target construct into other dimensions, they also impose construct-unrelated variance on the composite, further compromising its validity.

1.3.2. Limitations of existing approaches in identifying redundant and extraneous facets

A general, albeit less severe issue is that the existing scale construction approaches and strategies are aimed at item selection and evaluation, rather than at the facet level. Nonetheless, to various extents, item-selection procedures can be, and have been, applied to the evaluation and selection of facets (a detailed discussion and guidelines are presented in Smith et al., 2003). Also, more substantive approaches focused on the explication of facets and testing multi-faceted constructs have emerged within recent decades (Carver, 1989; Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Costa & McCrae, 1998; Hull, Lehn, & Tedlie, 1991). The key problem is that the existing approaches are not instrumental in identifying redundant facets and, to a lesser extent, extraneous facets, nor were they developed or intended for this purpose. Specifically, it is argued here that none of the approaches reliably disentangle specific variance and common variance, thus failing to detect all extraneous facets, whereas they are entirely unable to disentangle common variance and unique common variance, giving rise to redundant facets.

Although the rational-intuitive approach seems to encompass the largest number of specific methods (e.g., content analysis, focus groups, and evidence-oriented methods), coming up with an optimal representation of the construct based on theory and reasoning alone is virtually impossible. Items or facets that appear to be conceptually relevant may not represent variance attributable to the target construct. Furthermore, as discussed, even thematically and empirically related facets may not represent a unique aspect of the construct, relative to the other facets within the model.

The internal approach, which subsumes the variations of factor analysis, cannot identify redundant facets, because it targets the common variance and fails to show whether a facet occupies a unique part of the construct variance not already covered by one or more of the other facets. In fact, redundant facets are prone to have inflated factor loadings, leading to overrepresentations of certain manifestations of the construct and their variance within the total composite. Further, although this approach may reveal many extraneous facets, it cannot identify them reliably. Factor loadings depend on the hypothetical facets, the common variance of which is unlikely to represent the construct accurately. If a set of facets represents the construct poorly, extraneous facets are more likely to load on the latent composite. Also, extraneous facets are particularly likely to be retained where low cut-offs are used, which is a problem, given that there are no agreed-on criteria regarding factor loadings and communalities at which one should retain items in Exploratory Factor Analysis (Gignac, 2009).

In contrast to the internal approach, in which items or facets are selected based on their interrelationships, the external approach selects variables based on their ability to predict relevant external criteria. A variable's predictive ability has relevance for the identification of redundant and extraneous facets, as these should not occupy any unique variance linked to the target construct. However, the external approach is restricted to attributes for which individuals at the low, or high, extremes of a given characteristic can be somewhat objectively identified. Examples include extraverts and introverts, who are relatively easy to detect, or people suffering a particular disorder, based on diagnostic criteria. For many constructs, however, especially the fuzzy ones, there is little agreement as to what characterises people at the extremes, which relates back to the conceptual ambiguity of these constructs.

Smith et al. (2003) have discussed in considerable detail how these three item-selection procedures are not only applicable at the more substantive facet level but can also be enriched by means of incremental validity principles, with the aim of identifying and retaining uniquely predictive facets. However, due consideration to the criteria to be used for this purpose was not given. One major issue is that both facets and individual criteria comprise specific variance, unrelated to the construct one is aiming to operationalise. As a result, they can correlate due to sources other than the target construct. The issues involved in leveraging criteria for the purpose of assessing facets will be described in more detail within the next section.

1.4. Facet Benchmarking

1.4.1. Purpose and Description

The purpose of FB is to provide an instrument refinement method that will help advance the operationalisation and representation of psychological attributes, or the construct validity of measures, by identifying redundant and extraneous facets. It concerns the identification of these problem facets both within individual measures and across multiple measures. Unlike the other psychometric scale construction approaches, which operate at the item level (with implications for theory), FB primarily concerns the representation of the construct domain, or substantive validity. It is intended to supplement the existing psychometric approaches and, like the construct validation paradigm as a whole (in which it is situated), best viewed as an ongoing and indeterminate process. That is, for any given construct, repeated and varying applications across samples will increase certainty in the identification of redundant and extraneous facets.

Presently divided into five broad steps, FB seeks to uncover redundant and extraneous facets based on the principle that they should not occupy any unique variance of the construct, relative to other hypothetical facets. As discussed, the common, construct-based variance of redundant facets is already occupied by other facets, whereas extraneous facets do not overlap with the target construct. Consequently, both types of facet compromise, rather than enhance, the representation of the construct. Given its aims, FB also contributes to the construct homogeneity, or unidimensionality, of individual measures, a key principle that is increasingly emphasised in the literature (e.g., Simms & Watson, 2007; Smith & Zapolski, 2009). To quote Smith and Zapolski (2009), "the use of single scores to represent multidimensional processes cannot be defended" (p. 95).[1] Therefore, identification of redundant and extraneous facets does not merely help optimise criterion validity; it contributes to the overall construct validity, as evidenced by improvements in convergent and discriminant validity.

The premise of FB is that an alternative representation of a construct can be derived in a way other than using the measurement vehicles specifically developed to assess it. If such a variable can be obtained, it can be used as a benchmark to examine whether each of a construct's hypothetical facets occupies a unique part of the common variance. The problem is that individual criterion variables themselves are partial indicators of a construct that do not represent the construct variance accurately; there is no gold-standard criterion. Individual criteria that are theoretically influenced by the target construct and commonly used to assess its criterion validity are unlikely to qualify as a comprehensive representation (Cronbach & Meehl, 1955; Epstein, 1984; John & Benet-Martinez, 2000).

Another problem with individual criteria is that they are often multidimensional and cannot be expected to represent the construct variance exclusively (Smith & Zapolski, 2009). Due to any specific variance that these criteria could bring into the equation, there would be an increased chance of seeing predictive effects of extraneous facets and, to a lesser extent, redundant facets. Moreover, it is realistic that some facets correlate positively with a given criterion, while other facets of the same measure

[1] Practically, it is typical for the facets of a construct, or even the items of a facet, to exceed the construct boundaries. The key point is that the score used to represent a given construct is unidimensional, rather than representing multiple, weakly related or even entirely orthogonal dimensions.

correlate negatively with the same criterion (Matthias Ziegler, Danay, Schoelmreich, & Buehner, 2010). Using multiple analyses involving a diversity of individual criteria, therefore, would be no reasonable solution to representing the construct variance.

While using individual or multiple validation criteria is not instrumental for identifying redundant and extraneous facets, a single variable that is representative of the target construct's variance can be derived from the shared variance of a representative set of construct-relevant criteria, since these criteria are selected with the construct as a reference point. Using a latent composite of a balanced set of criteria appears to be a reasonable and practical solution to capturing the variance of a given construct accurately. Theoretically, such a variable can be considered an alternative representation of the construct; in practice, it may be generally best viewed as an approximation of its variance, with its accuracy depending on the method of derivation and knowledge about the construct already existing. Therefore, FB uses alternative representations of the construct variance, derived from construct-relevant criteria, as a benchmark to assess whether a facet occupies a unique portion of that variance (and ultimately qualifies as redundant or extraneous). The five-step process of FB is described next.

1.4.2. Application

1.4.2.1. Step 1: Derivation of representative sample of construct-relevant criteria

The first step of FB is to systematically derive a comprehensive sample of construct-relevant criteria to be administered along with a multi-faceted measurement instrument of the target construct to multiple samples. Given the aims of FB, it is imperative to stress the importance of ascertaining reliable measurement of *all* facets. Unreliability of facet measurement can attenuate the facets' contribution within a measure; profound differences in reliability compromise the facets' relative validity and bias the overall results they produce when examined simultaneously (Smith et al., 2003).

Step 1 requires determining and using an appropriate, systematic process of selecting construct-relevant criteria. The composite to be derived from these criteria is then used at Step 3 to assess whether each of the hypothetical facets occupies unique construct variance. The challenge is to select a set of criteria with shared variance that represents the construct variance comprehensively (i.e., not missing any parts) and exclusively (i.e., not imposing variance unrelated to the construct). Both these requirements inevitably involve a theoretical process (e.g., Ziegler et al., 2013), as is the case for facet selection. However, achieving an exclusive construct representation, or avoiding construct-unrelated variance, is considerably facilitated by the statistical procedure described in Step 2, possibly coupled with a more straightforward process surrounding criteria selection (compared to specifying facets). A criterion-based construct representation that extends beyond the boundaries of the construct, thus, is rather unlikely. The necessity of ascertaining a comprehensive representation can be met by repeated application of FB to the same set of facets (i.e., replication), but different, systematically selected sets of criteria for the purpose of deriving the criterion-based construct representation. Hereafter, the term *criterion-based composite* is used to refer to variables representing the shared variance of construct-relevant criteria.

As far as the theoretical component is concerned, different sampling procedures, or approaches to systematically deriving a representative set of criteria, are conceivable. The generally envisioned approach uses variables that are conceptualised as proximate (psychological) outcomes. These are variables representing affective, behavioural, cognitive processes that are directly linked to the construct and known to correlate in the expected direction with well-validated measures of the construct; more distant or indirectly-related criteria increase the chances of seeing unique effects of problem facets, as they may not represent the target construct primarily and, thus, diminish its representation accuracy. However, prior empirical correlations may not be necessary, and other, more theory-driven or qualitative approaches (Matthias Ziegler et al., 2015) may be incorporated in making these decisions. Another general principle that can be confidently stated is that the number of criteria necessary to represent a construct varies (positively) with the construct's breadth and level of abstraction.

It seems neither feasible nor necessary to identify and administer all relevant criteria, since many of them are likely to overlap in their common (construct) variance. Ideally, one would obtain a representative sample of all construct-relevant criteria without "duplicating" parts of the variance, thus aiming for a balanced representation; if the construct variance is not balanced in the in criterion-based composite, or if the constituent criteria are not balanced with respect to the construct variance, the common variance would shift towards individual facets, which then dominate. While it may not be entirely detrimental to have several unbalanced criterion-based composites, no single criterion-selection strategy may guarantee a comprehensive and accurate representation of the construct variance through the criteria. However, confidence in results will increase with repeated applications of FB to the same set of facets, but different sets of criteria, each derived systematically to aim for a comprehensive representation of the construct variance. Replication using not only the same, but also different sets of criteria is essential in order to identify any redundant or extraneous facets with confidence. Especially if it is uncertain whether a given set of criteria represents the construct accurately, repeated application of FB to multiple criterion-based composites, based on differing sets of criteria, is warranted—until it can be argued that the construct has been fully represented across studies.

Trait Activation Theory

One particular consideration that is warranted during the process of selecting criteria concerns situational moderators that may 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). Supporting this theory, research has shown that a given attribute can predict a certain criterion in some situations, but not in others (e.g., Barrick & Mount, 1991). Thus, if one were to sample very specific criteria from only some relevant situations, correlations of a valid facet with the criteria could be systematically negligible (in a worst-case scenario), rendering the facet redundant. It, therefore, is vital that the chosen criteria are either relevant across situations (i.e., general) or systematically sampled from all conceivable situations in which the construct manifests itself in some form.

1.4.2.2. Step 2: Extraction of criterion-based construct representation(s)

The second step is to conduct a Principal Component Analysis on the criteria administered to each sample in order to extract the first principal component, or criterion-based composite. Principal Component Analysis appears to be an appropriate extraction method for deriving the criterion-based composite, although similar results emerge for principal axis factoring in cases where commonalties are low (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Since principal components are linear and orthogonal composites of the measured variables, accounting for all of their variance, successive components account for variance not already explained by preceding components. In other words, where variables are theoretically derived to extract (as opposed to identify) a single dimension, the first component will generally show loadings of most (if not all) variables. By contrast, factors derived via principal axis factoring are intended to explain the shared variance of observed variables, which is not always best achieved by consistently high loadings on the first component. The advantage of principal component analysis, thus, is that the first principal component will generally yield a more accurate representation of the construct than the first factor obtained from principal axis factoring.

Theoretically, a disadvantage of principal component extraction is that it concerns the entire variance of among observed variables, including unique variance and error variance. However, this is not necessarily a limitation in the context of FB, since that variance will be subsumed under the last components extracted, or least likely under the first. In any case, principal component extraction seems to offer the better method than principal axis factoring for deriving the criterion-based composite.

Among all the empirical components that summarise the variance in these criteria, the first is, in theory, the variable that represents the construct (variance), because the criteria were selected using the construct as the reference point. Any unrelated criteria (i.e., those that do not load on the same component as the others) are identified and excluded in this process. FB can, thus, accommodate and, to some extent, resolve differences in how researchers define the target construct as well as in the criteria they consider relevant. Divergent criteria should exhibit low loadings on this first component, varying primarily due to sources other than the target construct. However, those criteria may still co-vary with the target construct, due to common-method effects (Campbell & Fiske, 1959) and other reasons. Consequently, they can introduce construct-unrelated variance on this component, which would, in turn, increase the chances of predictive effects for extraneous facets (or for the specific variance of redundant facets).

In general, it makes sense to proceed with a generic, minimum loading of .30, the common cut-off for scale items or facets. A pre-specified value is intended to foster reliability and replicability of results. However, it may be unwise to strictly advocate a specific cut-off, especially at this point, as elimination is a controversial issue that tends to elicit concerns by reviewers and editors (e.g., Ziegler, 2014). This minimum loading may change with further development of FB and, based on solid theoretical grounds, may even be adjusted between areas of application. For example, very narrow criteria that share relatively little variance with the construct are prone to be discarded at a cut-off of .30, but they may still occupy a unique part of the construct variance not already covered by other, possibly broader criteria. The important point is that the specification of this cut-off is made *a priori*, guided by strong reason and theory.

Factor (component) rotation seems inappropriate in the context of FB, since the aim is to ascertain loadings of theoretically relevant criteria on the criterion-based composite, not to examine the structure of latent factors and, thereby, disguise more accurate loadings.

1.4.2.3. Step 3: Identification of problem facets

Step 3 of FB examines whether each of the facets occupies a significant portion of variance in the derived criterion-based composite and if the variance explained is in the expected direction. Facets that fail to account for variance in this composite are likely to be redundant or extraneous and should be excluded from the set of facets used to represent the construct. The most straightforward statistical procedure for this purpose is to regress the criterion-based composite on the theoretical set of facets, using statistical regression (also referred to as the stepwise method), with all hypothetical facets entered at the initial step. Stepwise regression is the appropriate algorithm in this instance, as it both removes (criterion: $p \geq .05$) and possibly re-enters (criterion: $p < .05$) predictors one-by-one, based on their ability to account for unique variance in the criterion-based composite. Facets will be removed from the analysis successively if they do not explain unique variance in the criterion. In this process, the presence of redundant and extraneous facets may initially help suppress the (significant) effects of valid facets, although likely those with the least construct variance, contributing to their removal at initial steps. Yet, the stepwise method re-enters facets removed from the analysis at preceding steps if they gain their significant explanatory effect at later steps (i.e., upon removal of redundant and extraneous facets).[2] Of note, betas in an unexpected direction contradict how the facets concerned should behave theoretically and are detrimental to construct validity. If present at the final step, the analysis is to be repeated without such facets.

In sum, facets that always have non-significant beta weights across samples and construct representations are redundant or extraneous. To account for chance effects, facets that show significant betas on only rare occasions (e.g., less than 5% of the time) and of negligible magnitude may also be considered redundant or extraneous. In order to ascertain sufficient statistical power, the sample size should conform to accepted standards and best practices (Kelley & Maxwell, 2003; Tabachnick & Fidell, 2007),

[2] It is improbably that valid facets are excluded as a function of suppressor effects, which occur through combination of certain facets (e.g., redundant or extraneous facets) with any other facets. Once redundant and extraneous facets are removed, the valid facets should, by definition, regain their significant effect. To the contrary, redundant and extraneous facets should lose their significant effects sooner or later.

some of which revolve around the number of independent variables (predictors). On the other hand, research has shown that varying the sample size has relatively little influence on the number of predictors in automated subset selection algorithms (Derksen & Keselman, 2011).

Since the various automated selection algorithms have been heavily criticised (for good reasons), it seems imperative to justify the application of stepwise regression in this instance and explain why it does not constitute a limitation. A major concern is that these algorithms lack theoretical basis, operating purely on some pre-specified empirical criterion. In a typical research context, regression is used to answer various questions examining the predictors of an important criterion of interest. In those instances, it would indeed be unwise to rely on some automated selection procedure, especially when using different types of predictors. In the context of FB, however, the predictors are all of the same type (i.e., facets of the target construct) and the theoretical criterion of interest is an empirical one (whether the facets occupy unique construct variance); there is no theoretical order among the facets, and the focal question of unique common variance can only be answered statistically. Stepwise regression, as described above, can be used to address this question.

A second criticism concerns specifically the removal of predictors based on their ability to predict the criterion. High intercorrelations among predictors are generally considered problematic in multiple regression analysis, because they can compromise the explanatory effects of individual predictors (Pedhazur, 1997). However, in conjunction with the systematic removal of facets via stepwise regression, FB capitalises on this principle in order to identify redundant facets. Essentially, high intercorrelations mean that the predictors concerned are likely to explain (much of) the same variance in the criterion, rendering some as redundant. Regardless of their intercorrelations with other facets, extraneous facets should never explain any variance in alternative construct representations.

Another concern is that, given its multiple steps, the procedure is unduly influenced by chance features of the data and, hence, that the ensuing models are difficult to replicate. In testing multiple models, stepwise regression and is prone to overfitting the data. FB fully accounts for this limitation by means of built-in replication, conducted across the same and different criterion-based composites. However, it does not require the same solution to be obtained across samples. The crux are the predictors that never find their way into the regression models (final step of model), which renders FB a conservative, yet reliable, method for identifying problem facets. The problem of overfitting may be another desirable feature in the context of FB; it maximises the number of significant predictors and, thus, renders FB a conservative approach to detecting facets that fail to account for unique construct variance. In the context of FB, stepwise regression seems highly appropriate for the purpose of identifying facets that do not occupy a non-negligible part of the construct variance represented by the criterion-based composite. The bottom line is that the statistical limitations of stepwise regression are either not relevant to the questions FB seeks to address (e.g., elevated R^2 values, atheoretical), or they are compensated and justified by the design and particular focus of FB on non-significant predictors (e.g., replicability, chance effects, multicollinearity, narrow C.I.s, overfitting).

1.4.2.4. Step 4: Comparison of original and modified scale composites

The purpose of Step 4 is (a) to ascertain that no loss in construct validity occurred as a result of removing any problem facets and (b) to get a general idea of the magnitude of any improvement attained in construct validity. Hence, it also indicates what FB adds to the other approaches used in the development of psychometric measures.

A modified scale composite is computed from facets showing significant explanatory effects in at least one of the samples used, and then compared in its association with the criterion-based composite derived at Step 2 to that of the original scale composite. The criterion-based composite is used as a gauge for assessing the relative construct validity of the original and modified composites. Since a composite of all facets, including redundant and extraneous facets, averages predictive and non-predictive facets, its correlation with the criterion-based composite should in most cases be weaker than that of a modified composite encompassing predictive facets only (see Smith et al., 2003, for a more detailed discussion of this effect). As long as no loss occurs by their removal, evidenced by reduced associations with the criterion-based composites, non-predictive facets should be considered for exclusion from the measure. To compare correlations, Steiger's *Z* test can be performed, using Syntax or a more convenient external programme (FZT Computator; Garbin, n.d.).

Importantly, this step should not be regarded as the be-all-end-all gauge for assessing the level of improvement attained; optimising the validity of measures and, ultimately, advancing our understanding of the respective constructs is not the sole advantage of FB. Even if the gain in construct validity is minimal, there are benefits in terms of minimising the length of measures.

1.4.2.5. Step 5: Classification of non-predictive facets as redundant or extraneous

Step 5 serves to classify the identified problem facets as redundant versus extraneous. Their associations with the modified scale composite, and optionally also with the criterion-based composite, are examined. Consistently non-negligible associations will likely suggest that these facets are redundant, whereas non-significant associations suggest that they are extraneous.

1.5. Summary and Thesis Structure

Although proven tremendously useful, the established psychometric approaches are inadequate as far as the identification of problem facets is concerned. Specifically, they are unable to identify redundant facets, which lack unique common variance (relative to the other hypothetical facets), and extraneous facets, which have no common variance whatsoever as regards the target construct. Presumably, this limitation contributes to the diversification of measures and the plethora of facets seen in the psychometric literature of many constructs. It also compromises the validity of research findings based on psychometric measurement as well as assessment accuracy in psychological applications. For example, redundant and extraneous facets weaken the criterion validity of psychometric measures systematically.

An instrument refinement method, FB seeks to complement the existing psychometric approaches to scale construction, focusing specifically on the identification of redundant and extraneous facets. The method consists of a five-step process that sets out to establish whether each hypothetical facet occupies unique construct variance, relative to the other hypothetical facets of a given measure, a key requirement of any valid facet that has hitherto been neglected. Furthermore, FB classifies any problem facets as redundant versus extraneous, and ascertains that construct validity does not diminish as a result of excluding these facets (construct validity is theoretically always improved or maintained). Furthermore, FB is instrumental in identifying invalid content in a measure that increases administration time unnecessarily.

Three empirical chapters of the current thesis present applications and applications of FB, each in the context of a different construct. These chapters increase in their level of methodological rigour and scrutiny of the method, as explained in the respective parts of this thesis. **Chapter 2** is a preliminary application of FB, based on available data and applied to a measure of trait emotional intelligence. **Chapter 3** lays the psychometric groundwork for **Chapter 4**, which presents the first planned application of FB, centred on the construct of dispositional mindfulness. Mindfulness is currently one of the most heavily researched concepts in psychology and, for reasons to be specified, it is well-suited for the purpose of examining FB. **Chapter 5** presents the development and validation investigation of a measure of a novel construct, the General Factor of Motivation. The third application of FB presented in **Chapter 6** was applied to this measure. Finally, **Chapter 7** presents a summary of the findings concerning FB, as well as a discussion of their implications for research, theory, and applied psychological assessment. The wider implications of FB are also explored and avenues for future research on the method are proposed.

CHAPTER 2: Application of Facet Benchmarking in the Context of Trait Emotional Intelligence

2.1. Introduction

A construct of contemporary interest that illustrates the challenge of representing and operationalising constructs is emotional intelligence (EI). Much has been said about what constitutes EI, as is apparent from the diversity of EI models and operationalisations. The divergence of research into the two increasingly distinct subareas of trait EI and ability EI has brought some structure into the field. Petrides and Furnham (2001) pointed to the fundamentally distinct nature of constructs based on typical-performance, the predominant measurement method in the EI literature, as compared to those that are based on maximum-performance. But even when taking the split between typical- and maximum-performance measures into consideration, substantial discrepancies in how the construct is represented via structural models and arrays of facets remain across measures (cf. Dulewicz, Higgs, & Slaski, 2003; Jordan & Lawrence, 2009; Petrides & Furnham, 2001; Salovey, Mayer, Goldman, Turvey, & Palfai, 1995; Schutte et al., 1998; Tapia & Marsh, 2006; Tett, Fox, & Wang, 2005); the construct boundaries are far from agreed upon.

Trait EI has provided a framework for reconceptualising self-report measures of EI initially supposed to assess cognitive emotional abilities, which they are hardly able to measure (Freudenthaler & Neubauer, 2007; Paulhus, Lysy, & Yik, 1998). However, the distinction of ability and trait EI goes beyond mere operational differences in response format. For example, self-report measures based on Mayer and Salovey's (1997) four-branch ability EI model do not seem to measure trait EI comprehensively, as evidenced by their relatively weak construct validity compared to instruments developed to measure trait EI specifically (Gardner & Qualter, 2010; Martins, Ramalho, & Morin, 2010). By definition, trait EI refers to a compound trait located at the lower levels of personality hierarchies that integrates the affective aspects of personality

(Petrides, Pita, & Kokkinaki, 2007); it does not encompass emotion-related skills or abilities.

Trait EI is also conceptually distinct from the construct of social intelligence, irrespective of the method of measurement and conceptualisation of trait versus ability. Whereas the former concerns primarily emotional aspects of personality, the latter reflects how people interact with others (e.g., Petrides, Mason, & Sevdalis, 2011). Of course, this does not preclude overlap in their sets of facets, since many specific attributes integrate social and emotional qualities (e.g., aggression, assertiveness, and empathy). The key point is these abstract and difficult-to-define constructs are fundamentally distinct in their core. One would find considerably more emotional/affective facets within a measure of trait EI and more social/interpersonal facets in a measure of trait social intelligence.

This study will examine FB in the context of trait EI, as operationalised through the Trait Emotional Intelligence Questionnaire (TEIQue; Petrides, 2009). The TEIQue was designed to assess the construct of trait EI comprehensively and has hitherto produced very promising results in terms of construct validity (Freudenthaler, Neubauer, Gabler, Scherl, & Rindermann, 2008; Gardner & Qualter, 2010; Martins et al., 2010). Its theoretical set of 15 facets was determined through a content analysis of existing measures, retaining only those facets that were common across salient EI models. This unique approach captured the consensus among the existing models and measures, possibly yielding a more accurate representation of the target construct than other models. Evidence attesting that the TEIQue facets satisfy minimum standards for factor loadings has accumulated across translations of the measure (e.g., Freudenthaler et al., 2008; Martskvishvili, Arutinov, & Mestvirishvili, 2013; Mikolajczak, Luminet, Leroy, & Roy, 2007).

Although the model underlying the TEIQue has withstood the test of time, it is possible that some of the numerous facets on which it is based are redundant or extraneous. In this preliminary application of the FB, data gathered in previous psychometric studies of the TEIQue, including some of its translations (six samples in total), were used. The data from each sample included measurements of various construct-relevant criteria. This approach was deemed appropriate for this initial investigation, as the criteria assessed across these samples were diverse and representative of the four TEIQue factors. The principal components from the criteria assessed in each of the samples were extracted in order to provide alternative representations of trait EI (Step 2 of FB). These criterion-based composites were then regressed onto the 15 trait EI facets to identify any non-predictive facets (Step 3). A composite comprising facets with predictive effects in any one or more of the six samples was compared to the original 15-facet composite in terms of its associations with the six criterion-based composites (Step 4). Lastly, facets that did not occupy unique variance in any of the criterion-based composites were further examined to classify them as redundant versus extraneous (Step 5).

2.2. Method

2.2.1. Samples and criteria

The data came from five cross-sectional studies (six samples), in which the criterion validity of the TEIQue across different sets of criteria was investigated. The samples were selected based on their relevance to the present investigation, as their data comprised thematically related, proximate outcomes. Samples 1, 4, and 5 were Greek, Spanish, and Georgian, respectively, whereas Samples 2, 3, and 6 were British. The demographic characteristics of the six samples are summarised in Table 2.1. With the exception of Sample 5, additional details for the samples can be found in previously published studies (Gardner & Qualter, 2010; Petrides, Pérez-González, & Furnham, 2007; Petrides, Pita, et al., 2007).

Table 2.1. *Demographic Characteristics of Study Samples*

	Age (years)			Gender	
Sample (*N*)	*M*	*SD*	Range	Male	Female
1[a] (271)	25.47	5.88	19–56	92	179
2[b] (193)	22.83	6.16	18–60	74	118
3[b] (151)	22.01	6.07	19–54	30	121
4[c] (202)	23.16	3.35	18–45	35	167
5[d] (179)	25.58	13.73	17–74	60	117
6[e] (288)	36.45	11.78	18–79	67	221

Note. Samples 1, 4, and 5 were Greek, Spanish, and Georgian, respectively.
[a]Petrides, Pita, et al., 2007. [b]Petrides, Pérez-González, et al., 2007, Study 2. [c]Petrides, Pérez-González, et al., 2007, Study 3. [d]Martskvishvili et al., 2011. [e]Gardner & Qualter, 2010.

The criteria are presented in Table 2.2, together with their corresponding measures. These criteria are either entirely emotion-laden (e.g., depression, positive and negative affect) or integrate emotional and social aspects of functioning (e.g., aggression, coping styles, personality disorders, life satisfaction, alcohol-related problems, loneliness). Importantly, the criteria considered across all six samples represent each of the four TEIQue factors (Well-Being, Self-Control, Emotionality, and Sociability), as indicated in Table 2.2. Thus, they are suitable for deriving alternative representations of the trait EI variance, as required in Step 1 of FB.

Table 2.2. *Criteria and Measures Used across Study Samples*

	Variables	Measures	Trait EI factor represented
Sample 1	Life satisfaction	Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985)	WB
	Rumination	Emotion Control Questionnaire (Roger & Najarian, 1989)	SC, SOC
	Coping strategies	Coping Styles Questionnaire (Roger, Jarvis, & Najarian, 1993)	SC, EMO, SOC
Sample 2	Coping strategies	Coping Styles Questionnaire (Roger et al., 1993)	SC, EMO, SOC
	Depressive symptomatology	Center for Epidemiologic Studies Depression Scale (Radloff, 1977)	WB, EMO
	Depressogenic attitudes and beliefs	Dysfunctional Attitudes Scale (Weissman & Beck, 1978)	WB, EMO
Sample 3	Aggression types	Aggression Questionnaire (Buss & Perry, 1992)	SC, EMO, SOC
Sample 4	Positive and negative affectivity	Positive and Negative Affect Schedule (Sandín et al., 1999; Watson, Clark, & Tellegen, 1988)	WB, SC, EMO, SOC
	General depression	Beck Depression Inventory (2nd ed.; Beck, Steer, & Brown, 1996; Sanz, Navarro, & Vázquez, 2001)	WB, EMO
	Personality disorders	International Personality Disorder Examination (López-Ibor Aliño, Pérez Urdaníz, & Rubio Larrosa, 1996; Loranger, Janca, & Sartorius, 1997)	WB, SC, EMO, SOC
Sample 5	General depression	Beck Depression Inventory (1st ed.; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961)	WB, EMO
	State and trait anxiety	State-Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983)	WB, EMO, SOC
Sample 6	Aggression types	Aggression Questionnaire (Buss & Perry, 1992)	SC, EMO, SOC
	Social and emotional (family and romantic) loneliness	Social and Emotional Loneliness Scale for Adults–Short form (DiTommaso, Brannen, & Best, 2004)	EMO, SOC
	Eating-related problems	Eating Disorders Diagnostic Scale (Stice, Telch, & Rizvi, 2000)	WB, SC, EMO

Alcohol-related problems	Self-Administered Alcoholism Screening Test (Hurt, Morse, & Swenson, 1980)	WB, SC, EMO WB
Subjective happiness	Subjective Happiness Scale (Lyubomirsky & Lepper, 1999)	WB
Life satisfaction	Satisfaction with Life Scale (Diener et al., 1985)	

Note. Sample 1 measures were administered in Greek, Sample 4 measures in Spanish, and Sample 5 measures in Georgian. EI = emotional intelligence; WB = Well-Being; SC = Self-Control; EMO = Emotionality; SOC = Sociability.

2.2.2. Measures

All measures in this study were based on self-report, mostly using multiple-point response scales.

Trait EI

The full form of the TEIQue, which yields global, factor (4), and facet (15) scores, was administered to all six samples. Samples 1 to 4 completed the initial version (v. 1.00, 144 items), whereas Samples 5 and 6 completed the current version (v. 1.50, 153 items). Samples 2, 3, and 6 completed the TEIQue in its original language (English), whereas Greek, Spanish, and Georgian translations were administered to Samples 1, 4, and 5, respectively. The TEIQue was translated by the researchers who conducted the studies (Martskvishvili, Arutinov, & Mestvirishvili, 2013; Petrides, Pérez-González, et al., 2007; Petrides, Pita, et al., 2007).

The four factors and their constituent facets are Well-Being (self-esteem, trait happiness, and trait optimism), Self-Control (emotion regulation, stress management, and low impulsiveness), Emotionality (emotion perception, trait empathy, emotion expression, and relationships), and Sociability (assertiveness, emotion management, and social awareness). Two facets (adaptability and self-motivation) have not been included in any of the four factors, but contribute directly to the global score. More detailed descriptions of the facets and factors can be found in Petrides (2009). The TEIQue items are responded to on a 7-point Likert scale, ranging from 1 (*disagree completely*) to 7 (*agree completely*). Internal consistencies at the facet level were predominantly within a range of .70 to .80 across studies. Cronbach’s alphas for global trait EI ranged from .81 (Sample 5) to .96 (Sample 6).

Criteria

A summary of the criterion measures and references can be found in Table 2.2. The measures administered to Sample 1 were translated by the authors who conducted the study. For Samples 4 and 5, the criteria were assessed with available translations of the measures.

Sample 1. The Satisfaction with Life Scale (Diener et al., 1985) consists of five items that yield a global life satisfaction score (e.g., "In most ways my life is close to my ideal") measured on a 7-point Likert scale. Cronbach's alpha in this sample was .84.

The 14-item rehearsal subscale from the Emotion Control Questionnaire (Roger & Najarian, 1989) was used as a measure of rumination (e.g., "I remember things that upset me or make me angry for a long time afterwards"). Items are responded to on a 7-point Likert scale. Cronbach's alpha was .84.

The Coping Styles Questionnaire (Roger et al., 1993) consists of 60 items assessing four coping strategies. Two of these (rational and detached coping) are considered to be adaptive, and the other two (emotional and avoidant coping) maladaptive. Items are responded to on a 4-point Likert scale. Cronbach's alphas were .81 (rational coping), .80 (detached coping), .84 (emotional coping), and .68 (avoidant coping).

Sample 2. Sample 1 completed a Greek translation of the Coping Styles Questionnaire, while Sample 2 completed the original English version. Cronbach's alphas were .82 (rational coping), .84 (detached coping), .83 (emotional coping), and .68 (avoidant coping).

The Center for Epidemiologic Studies Depression Scale (Radloff, 1977) is a 20-item measure of depressive symptomatology, specifically developed for use in non-clinical settings. Respondents indicate how frequently they experience a range of depressive symptoms during the past week (e.g., “I was bothered by things that usually don’t bother me”). Items are responded to on a 4-point Likert scale. Cronbach’s alpha was .92.

The Dysfunctional Attitudes Scale (Weissman & Beck, 1978) is a measure of depressogenic attitudes and beliefs, based on a cognitive theory perspective and consisting of two parallel 40-item forms. Using a 7-point Likert scale, respondents answer each item according to how they think most of the time. Form A was administered to Sample 2, yielding an alpha level of .87.

Sample 3. The Aggression Questionnaire (Buss & Perry, 1992) assesses four distinct types of aggression. It consists of 29 items responded to on a 5-point Likert scale. The four aggression scales, and their respective internal consistencies, are physical aggression (.80), verbal aggression (.69), anger (.80), and hostility (.79).

Sample 4. The Positive and Negative Affect Schedule (Sandín et al., 1999; Watson et al., 1988) was used to assess positive and negative affect. Each affective dimension has 10 items that are responded to on a 5-point Likert scale. The alpha level was .89 for positive affect and .85 for negative affect.

The second edition of the Beck Depression Inventory (Beck et al., 1996; Sanz et al., 2003) was administered to this sample. It measures the severity of depression and consists of 21 items that are responded to on a 4-point Likert scale. The alpha level was .87.

The International Personality Disorder Examination (López-Ibor Aliño et al., 1996; Loranger et al., 1997) has a semi-structured interview format aligned to the ICD-10 and DSM-IV criteria. Typically used as a screener, this instrument comprises 77 dichotomous true-or-false items that produce scores representative of 10 distinct personality disorders. Alpha levels were generally low to moderate, ranging from .32 for Schizoid to .67 for Avoidant.

Sample 5. The first edition of the Beck Depression Inventory (Beck et al., 1961) was administered to Sample 5. Like its successor, which was administered to Sample 4, this edition measures the severity of depression and consists of 21 items that are responded to on a 4-point Likert scale. The alpha level was .81.

The State-Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) comprises 40 items, which are based on a 4-point Likert scale and represent two types of anxiety: state and trait anxiety. Accordingly, scores can be derived for both state and trait anxiety, which had alpha levels of .85 and .81, respectively.

Sample 6. The Aggression Questionnaire (Buss & Perry, 1992), as described in Sample 3, was also administered to this sample. The internal consistencies were .71 for physical aggression, .65 for verbal aggression, .66 for anger, and .69 for hostility.

The Social and Emotional Loneliness Scale for Adults–Short Form (DiTommaso et al., 2004) contains 15 items that are responded to on a 7-point Likert scale. The items are evenly distributed across three subscales assessing family loneliness ($\alpha = .89$), romantic loneliness ($\alpha = .96$), and social loneliness ($\alpha = .89$).

The Eating Disorders Diagnostic Scale (Stice et al., 2000) consists of 22 items, 19 which (items 1–18 and 21) are used to derive the single composite of this scale. One of the 19 items (item 21, addressing amenorrhea) was omitted in order to make the scale suitable for participants of both genders. The measure's items have a mix of Likert-type and *yes*-or-*no* response formats. In this sample, the internal consistency was .86.

The Self-Administered Alcoholism Screening Test (Hurt et al., 1980) consists of 35 dichotomous *yes*-or-*no* items, indicative of alcohol-related problems. Its internal consistency in this sample was .76.

The Subjective Happiness Scale (Lyubomirsky & Lepper, 1999) consists of four items that are responded to on a 7-point Likert scale. Its internal consistency in this sample was .89.

The Satisfaction with Life Scale (Diener et al., 1985) previously described in Sample 1 was also administered to this sample, in which it had an alpha level of .90.

2.2.3. Statistical analyses

The criteria corresponding to each sample were submitted to a Principal Component Analysis to derive the criterion-based composites. To ascertain that the first component of each, presumably multidimensional, set of criteria represents the target construct, rather than any other dimension, a rather strict cut-off was used (also in view of the overall diversity of criteria across samples, none of which were specifically tested to examine FB). Specifically, criteria were included within the respective criterion-based composite if they had loadings either (a) of at least .50, or (b) or of .30–.49 that were greater than their loadings on ensuing components. These variables were deemed to be too distinct from the target construct, with additional dimensions implicit in them increasing the chances of predictive effects for extraneous facets (or for the specific variance of redundant facets).

The derived criterion-based composites were regressed onto the 15 trait EI facets, using the stepwise method in each analysis. All facets were entered at the first step and then removed successively, starting with the least significant one. Since the stepwise method was used, as specified by FB, it was possible for facets already removed to be re-entered at later steps of the analyses.

The original composite of all 15 trait EI facets and a composite comprising facets included in the final model in at least one of the six regression analyses were compared in terms of their associations with the criterion-based composites. One reason for including facets with significant predictive effects in any of the six samples in this composite is to account for variations in the criteria used to derive the criterion-based composites. Steiger's *Z* tests were computed to examine if there are significant differences in the correlations of these two composites with the criterion-based composites across samples.

To differentiate between redundant and extraneous facets, zero-order correlations of any non-predictive facets with a modified composite comprising the predictive facets only were also examined. In theory, redundant facets should correlate significantly with the global construct, whereas extraneous facets should show correlations closer to zero.

2.3. Results

2.3.1. Dimensional reduction of criteria

Results of the Principal Component Analyses for the criteria used in each sample are presented in Table 2.3. The only variable excluded from Samples 1 and 2 was avoidance coping because it had relatively weak loadings (.14 and -.46, respectively) on the first principal component. It also resulted in bifactorial solutions in the initial analyses, loading considerably higher on the second component. For the same reasons, three personality disorders were removed from the final analysis of the Sample 4 criteria: schizoid, histrionic, and narcissistic. Their respective loadings on the first principal component were .38, .36, and .24, and lower than their loadings on a second or third component. Two variables, verbal aggression and eating-related problems, were excluded from the Sample 6 criteria. Their loadings on the first principal component were .32 and .27, respectively, and both loaded much higher on additional components. These seven variables were excluded on the grounds that they were too different from the target construct. With these variables omitted, a latent composite was derived from the remaining variables in Samples 1, 2, 4, and 6. All criteria assessed in Samples 3 and 5 were included in their respective composites, as they all loaded highly on a single principal component.

Table 2.3. *First Principal Component Loadings for Criteria in Each Sample*

	Variable	Factor loading	Communality	% of variance
Sample 1	Life satisfaction	.63	.40	51.87
	Rumination	.59	.35	
	Rational coping	.78	.61	
	Detached coping	.80	.64	
	Emotional coping	-.77	.59	
Sample 2	Rational coping	.77	.59	55.37
	Detached coping	.77	.59	
	Emotional coping	-.83	.70	
	Depressogenic attitudes and beliefs	.55	.30	
	Depressive symptomatology	.77	.59	
Sample 3	Physical aggression	.73	.53	52.39
	Verbal aggression	.63	.39	
	Anger	.86	.73	
	Hostility	.66	.44	
Sample 4	IPDE paranoid	.73	.58	44.42
	IPDE schizotypal	.76	.62	
	IPDE antisocial	.52	.62	
	IPDE borderline	.78	.61	

	IPDE obsessive-compulsive	.48	.32	
	IPDE dependent	.58	.41	
	IPDE avoidant	.68	.47	
	Negative affect	.73	.54	
	Positive affect	-.53	.61	
	General depression	.78	.65	
Sample 5	Depression	.83	.68	74.42
	State anxiety	.89	.79	
	Trait anxiety	.87	.76	
Sample 6	Physical aggression	.44	.61	40.53
	Anger	.53	.71	
	Hostility	.75	.61	
	Social loneliness	.62	.52	
	Family loneliness	.63	.56	
	Romantic loneliness	.58	.45	
	Alcohol-related problems	.37	.23	
	Subjective happiness	-.80	.65	
	Life satisfaction	-.83	.72	

Note. Avoidance coping was excluded from Samples 1 and 2, as it loaded relatively weakly on the first principal component and more strongly on a second component. For the same reason, the IPDE schizoid, histrionic, and narcissistic scales were excluded from Sample 4, and verbal aggression and eating-related problems from Sample 6. IPDE = International Personality Disorder Examination (Loranger et al., 1997).

2.3.2. Regression of criterion-based composites on facets

Summaries of the stepwise regression analyses with the criterion-based composites as the dependent variables are presented in Table 2.4. Due to the large amount of data, only results for the initial and final models as well as beta weights for facets retained in the final model are presented. While all 15 facets were initially included in the analyses, facets that were not retained in the last step of any of the six regression models are omitted from Table 2.4. The analyses for Samples 3, 4, and 6 excluded the facet of emotion management, while that for Sample 6 additionally excluded the facets of trait empathy and emotion perception. The reason for omitting these facets is that, when initially included, the direction of their explanatory effect was opposite to those of the other facets in the equations.

Table 2.4. *Stepwise Regression Analysis Summaries for TEIQue Facets Predicting the Criterion-Based Composites*

	Sample 1		Sample 2		Sample 3		Sample 4		Sample 5		Sample 6	
Trait EI facets	β	R^2_{Adj}	β	R^2_{Adj}	β	R^2_{Adj}	β	R^2_{Adj}	β	R^2_{Adj}	β	R^2_{Adj}
Model 1 (all facets)		.68		.72		.37		.59		.54		.77
Final model		.67		.72		.38		.58		.54		.76
Self-motivation											-.10**	
Emotion regulation	-.20***		-.20***		-.21**							
Trait happiness					-.24**		-.29***		-.25***		-.58***	
Low impulsiveness					-.19*		-.12*				-.11**	
Self-esteem	-.20***		-.20**						-.31***			
Assertiveness			-.14**				-.17***					
Trait optimism	-.30***		-.27***									
Relationships					-.26***		-.21***				-.21***	
Adaptability	-.12*						-.13*					
Stress management	-.24***		-.33***				-.18**		-.36***		-.10**	
ΔR^2		-.02		-.01		-.03		-.02		-.01		-.00
N		271		193		151		202		179		288

Note. Only beta weights for facets retained in the final models are displayed. EI = emotional intelligence. TEIQue = Trait Emotional Intelligence Questionnaire (Petrides, 2009).
*$p < .05$. **$p < .01$. ***$p < .001$.

Of the 15 trait EI facets, five did not explain unique variance in the criterion-based composites in any sample and, thus, do not appear in the final regression models. These facets were trait empathy, emotion perception, emotion expression, emotion management, and social awareness. In addition to being manually excluded from Samples 3, 4, and 6, emotion management did not appear in the final regression models in Samples 1, 2, and 5, based on the stepwise method. Likewise, trait empathy and emotion perception, which were manually removed from the Sample 6 regression, were non-predictive in the other samples. Therefore, neither these three facets nor the two non-predictive facets appear in Table 2.4. Of the 10 facets showing significant predictive effects, one (stress management) accounted for unique variance in five samples, one (trait happiness) accounted for unique variance in four samples, four (emotion regulation, self-esteem, low impulsiveness, and relationships) accounted for unique variance in three samples, two accounted for unique variance in two samples (assertiveness and trait optimism), and two, self-motivation and adaptability, accounted for unique variance in one sample.

In comparing the additive predictive effects of all 15 facets included in the initial prediction model (shown as Model 1) against the final set of facets remaining in the last step of each regression analysis (shown as Final model), the appropriate statistic to examine is the adjusted R^2, which can account for the unequal degrees of freedom. As is apparent across all six samples, the shortened sets accounted for virtually the same amount of the variance as the 15-facet composite. Even the unadjusted change in R^2 from the initial to final model was negligible and non-significant in the six samples. As discussed, however, regression analysis does not reveal the impact of non-predictive facets or facets with atheoretical, inverted effects on the explanatory power of higher-order composites, such as global trait EI. For example, the non-predictive facets of emotion expression and trait empathy can be expected to weaken the convergence of global trait EI with the criterion-based composites, since they are averaged along with the predictive facets within the global trait EI score. Hence, two trait EI composites comprising 15 and 10 facets, respectively, were compared in terms of their associations with the criterion-based composites.

2.3.3. Correlations of original and modified scale composites with the criterion-based composites

Pearson correlations of the 15- and 10-facet trait EI composites with the criterion-based composites are presented in Table 2.5. Also shown are Steiger *Z* tests of significant differences in the convergent validity of the two composites. Except for the latent composite derived from the Sample 3 criteria, associations of both trait EI composites with the criterion-based composites were strong across samples. Unlike the other samples, in which a latent composite of more diverse emotion-related criteria was used, the criterion-based composite derived from the aggression variables in Sample 3 was fairly homogenous and narrow, and, thus, least representative of trait EI. Correlations of the 10-facet composite with the criterion-based composites were larger than those of the 15-facet composite. In fact, the Steiger *Z* results indicate that the 10-facet composite had significantly greater convergent validity in all six samples.

Table 2.5. *Correlations of the Original and Modified TEIQue Composites with the Criterion-Based Composites*

Sample (*N*)	Original scale composite	Modified scale composite	Steiger's *Z*
1 (271)	.73	.79	4.94**
2 (193)	-.75	-.80	3.88**
3 (151)	-.49	-.58	3.79**
4 (202)	-.73	-.76	2.34*
5 (179)	-.65	-.68	2.27*
6 (288)	-.78	-.81	3.10**

Note. All correlations are significant at $p < .001$. "Original" refers to the scale composite of all 15 facets; "Modified" refers to the 10-facet scale composite minus the five non-predictive facets. TEIQue = Trait Emotional Intelligence Questionnaire (Petrides, 2009).
*$p < .05$. **$p < .01$.

2.3.4. Correlations of non-predictive facets with the modified scale composite

Correlations between the five non-predictive facets and the 10-facet composite are shown in Table 2.6. All correlations were significant and all except one (emotion management in Sample 3) were within a moderate range of .3 to .7, indicating that the facets are redundant, rather than extraneous.

Table 2.6. *Correlations of the Five Non-Predictive Facets with the Modified TEIQue Composite*

Sample (*N*)	Trait empathy	Emotion perception	Emotion expression	Emotion management	Social awareness
1 (271)	.32	.51	.38	.43	.66
2 (193)	.34	.48	.52	.46	.70
3 (151)	.35	.49	.50	.21*	.63
4 (202)	.46	.57	.40	.32	.64
5 (179)	.36	.52	.44	.36	.54
6 (288)	.36	.47	.48	.32	.57

Note. Correlations not denoted by an asterisk are significant at $p < .001$. TEIQue = Trait Emotional Intelligence Questionnaire (Petrides, 2009).
*$p < .01$.

2.4. Discussion

Application of FB to trait EI data from six European samples yielded promising results. Five facets did not explain unique variance in alternative representations of the construct variance, derived from varying sets of validation criteria. Removal of these five facets from the global trait EI composite significantly improved its associations with the criterion-based composites in all samples. Collectively, the results suggest that the five non-predictive facets overlap entirely with the predictive facets in their reliable common variance (i.e., variance attributed to the construct of trait EI), apparently compromising the construct validity of the global trait EI composite. It seems that the modified 10-facet composite gives a better representation of trait EI than the original composite.

The trait EI facets identified as non-predictive came exclusively from the TEIQue factors of Emotionality and Sociability. Notably, these two factors have shown little success in explaining incremental criterion variance vis-à-vis the other factors in previous research (Mikolajczak, Luminet, & Menil, 2006; Mikolajczak et al., 2007; Mikolajczak, Roy, Verstrynge, & Luminet, 2009; Siegling, Vesely, Petrides, & Saklofske, accepted; Swami, Begum, & Petrides, 2010; Uva et al., 2010). In only one study, one of these two factors (Sociability) accounted for incremental criterion variance, predicting somatic symptoms amid stress over mental and physical status, together with the Self-Control factor (Mikolajczak et al., 2006). However, it is important to remember that individual criteria are unlikely to represent the variance of the target construct very well and, therefore, significant predictive effects of redundant and extraneous elements are possible.

While all facets incorporated under the Self-Control and Well-Being factors explained incremental variance in the expected direction in at least one of the samples of the present study, the Sociability and Emotionality factors had only a single facet each that occupied variance in at least one of the criterion-based composites. Zero-order correlations of the non-predictive facets with the 10-facet composite were within a moderate range and significant, suggesting that the identified facets are redundant.

2.4.1. Implications

A shared characteristic of the five non-predictive facets is their integration of interpersonal emotional attributes, although some merge interpersonal and intrapersonal qualities (e.g., emotion perception represents the propensity to perceive emotions in oneself and in others). This pattern is in accordance with some evidence speaking to the distinctiveness of these types of facets (Siegling, Saklofske, Vesely, & Nordstokke, 2012; Siegling, Vesely, & Saklofske, 2013). As discussed previously (Siegling et al., 2012, 2013), it is possible that some of these facets (e.g., emotion management of others, trait empathy) share most of their variance with constructs more indicative of social behaviour, such as trait social intelligence (Petrides et al., 2011).

Empirical characteristics of redundant and extraneous facets are failing to occupy unique construct variance and compromising the construct validity of the global composite. Redundant facets share the same common variance with one or more of the other facets, giving disproportional weight to particular segments of the construct variance. Extraneous facets lie wholly beyond the target construct's boundaries, thus lacking common variance (i.e. their variance is due to dimensions other than the one targeted). Therefore, neither of these types of facet seems to take up unique common variance, thus weakening the construct validity of the model that incorporates them and of its operational vehicles. Overall, the results provide preliminary evidence for the efficacy of FB in identifying redundant facets, since all of the non-predictive facets seemed to fall into this category. At least in theory, it should also screen out facets that are completely extraneous and somehow found their way into the researcher's model.

2.4.2. Limitations and future directions

Although a similar set of predictive facets is likely to emerge in independent samples and across different criterion-based composites, fluctuations in terms of which facets will have significant effects are still possible. A statistical factor to consider is that facets may emerge as significant or non-significant due to chance. Self-motivation may be such a candidate, as it had a significant incremental effect in only one of the six samples, and the regression weight for its effect was very small. Although a scenario of all five (presumably redundant) facets being unrepresented in the criteria is highly unlikely, it is also possible that some segments of the construct variance were not represented in the criteria that were investigated. Consequently, facets related to any underrepresented construct variance would not have reached significance. While large fluctuations in the pattern of predictive facets are not expected, repeated applications of FB to TEIQue data are encouraged to increase confidence in the present findings. It is important to cross-validate the results in independent samples and sets of criteria that have not been previously used.

Further validation of FB with respect to other personality constructs is needed to provide definitive evidence for its efficacy. Whereas this chapter presents the initial application of the proposed, based on existing data, future studies designed specifically for its evaluation can yield more conclusive results. However, this is not to undermine the utility and relevance of using existing datasets, as FB requires evidence from numerous and relatively large samples. Applications of the analyses performed here by others who have suitable data (ideally, from multiple samples) are highly encouraged.

In designing future studies specifically for applying FB, it will be important to sample systematically from the entire theoretical range of relevant criteria to represent the variance of the target construct as comprehensively as possible. A second question to be addressed in further validation studies of FB is whether using the same measurement format for all variables introduces confounding effects in favour of FB. Measuring the criteria in the same way as the hypothetical facets creates common-method variance, and therefore common-method effects, such as socially desirable responding (Bäckström, Björklund, & Larsson, 2009; M. Ziegler & Buehner, 2009), may contribute to the pattern of results. Although is not particularly well understood how most of these method effects impinge on validity (Matthias Ziegler et al., 2013), the use of alternative methods (i.e., other than self-report) for assessing criteria relevant to trait EI and other personality constructs (e.g., informant ratings, behavioural observations, electronic diaries, and possibly biodata) can alleviate concerns regarding method variance. Converging evidence from applications of FB across criterion-based composites will eventually help us arrive at a consensus regarding the best set of facets for representing established, yet still partially elusive, individual-differences constructs.

2.5. Conclusions

Subject to further validation, FB seems to have utility in optimising multi-faceted assessment instruments. As discussed, a unique strength of the proposed strategy lies in its potential to identify redundant or extraneous facets, which conventional approaches do not accomplish. If validated rigorously and in the context of additional constructs, the method may have much to add to the construction of psychological assessment instruments, with possibly far-reaching implications for research and applied psychological assessment.

CHAPTER 3: Dispositional Mindfulness: Addressing Basic Psychometric Issues

3.1. Introduction

Another relatively recent and fuzzy construct that lends itself to the application of FB is mindfulness. Mindfulness, which can be very broadly understood as living in, and accepting, the present moment non-judgementally, as opposed to being preoccupied (Brown, Ryan, & Creswell, 2007; Heidenreich, Ströhle, & Michalak, 2006; Sauer et al., 2012), has generated a great deal of interest in applied and academic psychology. In applied psychology, it has led new approaches to treating mental illness and developing well-being (Kabat-Zinn, 1994; Segal, Teasdale, & Williams, 2002). In academic psychology, the concept has extended beyond its clinical applications to a focus on individual differences. This interest is evident in the recent spurt in psychometric research and proliferation of scales occurring in the past 10 to 15 years, most of which focuses dispositional, or trait, mindfulness (average or baseline states of mindfulness), rather than state mindfulness, or the particular mindful state at the time of measurement (Bergomi, Tschacher, & Kupper, 2013; Park, Reilly-Spong, & Gross, 2013; Sauer et al., 2012).

Research findings have been promising, with existing measures of mindfulness predicting criteria such as emotion dysregulation (Vujanovic, Bonn-Miller, Bernstein, McKee, & Zvolensky, 2010), sexual body esteem (Fink, Foran, Sweeney, & O’Hea, 2009), insomnia (Ong, Shapiro, & Manber, 2009), nicotine dependence and withdrawal (Vidrine et al., 2009), as well as relationship satisfaction and stress (Barnes, Brown, Krusemark, Campbell, & Rogge, 2007). However, as is typical at a relatively early stage in the research history of a concept (Furnham, 1990), disagreement spanning the operationalisation and, to a lesser extent, conceptualisation of mindfulness characterises the existing literature. As discussed elsewhere (Bergomi et al., 2013; Hart, Ivtzan, & Hart, 2013; Sauer et al., 2012) and described below, the set of mindfulness scales that has emerged can be described as heterogeneous, especially in terms of domain representation.

To streamline the application of FB to mindfulness measures (**Chapter 6**), the investigation reported in the current chapter had two general aims. The first was to assess the homogeneity (heterogeneity) of the various scales, by examining and, to some extent, cross-validating their convergent validity, shared underlying dimensions, and linkages to the Five-Factor Model (FFM; Study 1). The second aim was to examine the similarities, validity, and dimensionality of mindfulness facets and subscales across three independently developed measures (Study 2). As discussed, FB requires that the measures to which it is applied measure single construct, rather than multiple weakly related dimensions. The present investigation sought to establish that this requirement is met.

3.1.1. Measures and facets of dispositional mindfulness

Eight measures have been salient in the literature (Bergomi et al., 2013), although newer ones are emerging (Bergomi, Tschacher, & Kupper, 2012; Erisman & Roemer, 2011). The unidimensional facet or item measures organise their facets or items within a hierarchical model, under a single mindfulness factor (detailed descriptions can be found in Bergomi et al., 2013; Hart, Ivtzan, & Hart, 2013; Park et al., 2013; Sauer et al., 2012). While these measures may all represent the same construct, they are diverse in terms of their underlying structural models and representations of mindfulness; some of the measures are broader in scope, presumably assessing the construct more comprehensively, whereas others have a narrower focus, measuring only some of its elements.

The Five Facet Mindfulness Questionnaire (FFMQ; Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006), the Kentucky Inventory of Mindfulness Skills (KIMS; Baer et al., 2004), the Southampton Mindfulness Questionnaire (SMQ; Chadwick et al., 2008), and the Cognitive and Affective Mindfulness Scale-Revised (CAMS–R) comprise either four or five facets that vary between measures. However, only the FFMQ and KIMS have facet scores suitable for use in research and of satisfactory reliability (Baer et al., 2004, 2006), while the SMQ and CAMS–R use facets for representational purposes only (that is to say, the content of the facets is represented in a total score, but the measures do not yield facet scores per se; Chadwick et al., 2008; Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2006). Another two unidimensional measures, the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003) and the Freiburg Mindfulness Inventory (FMI; Walach, Buchheld, Buttenmüller, Kleinknecht, & Schmidt, 2006) directly operationalise the general mindfulness factor from their respective items; they do not use facets to represent the construct.

A distinct measure based on a hierarchical model is the Langer Mindfulness Scale (LMS; Bodner & Langer, 2001; Pirson, Langer, Bodner, & Zilcha, 2012), which is grounded in a somewhat divergent conceptualisation of mindfulness: as "a state in which one is open to novelty, alert to distinctions, sensitive to context, aware of multiple perspectives, and oriented in the present" (Bodner & Langer, 2001, p. 1). Its facets are Novelty Seeking, Novelty Producing, and Engagement. The underlying conceptualisation has been described as a Western approach that, despite similarities, differs from the traditional perspectives, which are linked to Eastern religions and provide the basis for the bulk of psychometric measures. More detailed information on differences between conceptualisations can be found in published reviews, such as in Hart et al. (2013).

The FFMQ was empirically derived by factor-analysing the items of the other five unidimensional facet or item scales (KIMS, CAMS–R, SMQ, MAAS, and FMI). For this reason, it can be considered a relatively comprehensive operationalisation of the construct that may supersede its constituent scales in terms of construct validity. The FFMQ consists of five facets (Describe, Act with Awareness, Accept without Judgment, and Nonreact), four of which (not including Nonreact) also constitute its main predecessor, the KIMS (Baer et al., 2004).

On the other hand, the FFMQ model, and in particular its Observe facet, has produced some problematic results. While intercorrelations among the FFMQ facets are generally significant and weak-to-moderate, as one would expect, the Observe facet has often shown non-significant, and sometimes even negative, correlations with one or more of the other four facets, such as Act with Awareness and Accept without Judgment, as well as weak factor loadings (Baer et al., 2004, 2006; Bohlmeijer, ten Klooster, Fledderus, Veehof, & Baer, 2011; Curtiss & Klemanski, 2014; Höfling, Ströhle, Michalak, & Heidenreich, 2011; Tran, Glück, & Nader, 2013). Moreover, a four-factor hierarchical model omitting the Observe facet tends to results in better model fit for the FFMQ than the originally envisaged five-factor model; increasing evidence supports a five-factor structure including Observe in meditators only (Aguado et al., 2015; Baer et al., 2004, 2006; Cebolla et al., 2012; Christopher, Neuser, Michael, & Baitmangalkar, 2012; Curtiss & Klemanski, 2014; Williams, Dalgleish, Karl, & Kuyken, 2014). Alternatively, a bidimensional facet model incorporating all five facets under two weakly associated second-order factors has also been identified and partially confirmed for a short form of the FFMQ in both meditators and non-meditators (Tran et al., 2014, 2013). In terms of criterion validity, FFMQ Observe was found to buffer the effect of stress in meditators only (Neale-Lorello & Haaga, 2015) and to have negligible incremental validity over the other facets in predicting construct-relevant criteria, including some detrimental effects (Cash & Whittingham, 2010; Christopher & Gilbert, 2009; Consedine & Butler, 2014; Vujanovic et al., 2010).

A further two measures are grounded in the mainstream conceptualisation of mindfulness, but they diverge operationally in their bidimensional structure, consisting of two subscales that correlate weakly or non-significantly. These are the Toronto Mindfulness Scale (TMS; Davis, Lau, & Cairns, 2009; Lau et al., 2006) and the Philadelphia Mindfulness Scale (PHLMS; Cardaciotto, Herbert, Forman, Moitra, & Farrow, 2008). The PHLMS was explicitly designed to operationalise two orthogonal subscales, labelled Awareness and Acceptance, which did not correlate (r = -.06; Cardaciotto et al., 2008). Although the TMS was created to permit oblique factors, its subscale correlations were not large enough to argue that a single shared dimension accounts for much of their variance, and they were only reported for the state version (r = .26 to .42; Lau et al., 2006). Thus, its two subscales were interpreted as assessing distinct, but related, latent constructs, labelled Curiosity and Decenter. It is important to bear in mind that using heterogeneous measures consisting of weakly related or orthogonal factors to represent a single construct is problematic (Smith, McCarthy, & Zapolski, 2009; Smith & Zapolski, 2009). Although neither of these two measures claims to assess a single global construct, both are linked to the extant literature (i.e., the concept of mindfulness) and depart from the other measures in their bidimensional focus.

Table 3.1 presents definitions for the FFMQ facets (Describe, Act with Awareness, Accept without Judgment, and Nonreact) and PHLMS and TMS subscales, along with sample items. A triplet of similar facets across the three focal measures consists of FFMQ Observe, PHLMS Awareness, and TMS Curiosity. Despite some differences, all three concern a deliberate perceptual focus on present-moment experiences. A pair of very similar facets consists of FFMQ Accept without Judgment and PHLMS Acceptance, both of which reflect a person's tendency to accept, rather than judge, internal and external experiences. Another pair of similar facets is that of FFMQ Nonreact and TMS Decenter, both reflecting (emotional) disengagement from one's inner feelings, perceptions, and thoughts.

Table 3.1. *Operationalisation of Mindfulness across the FFMQ, PHLMS, and TMS, Including Facet Definitions and Sample Items*

Measure and facets	Definition	Sample item
FFMQ		
Observe*	Tendency to observe, notice, or attend to internal and external phenomena.	I intentionally stay aware of my feelings.
Describe	Tendency to Describe or label sensations, perceptions, thoughts, emotions, etc. with words.	My natural tendency is to put my experiences into words.
Act with Awareness	Tendency to focus undivided attention on the current activity or avoiding automatic pilot; concentration.	I easily get lost in my thoughts and feelings.
Accept w/o Judgment**	Tendency to accept without making judgements or evaluations.	I disapprove of myself when I have irrational ideas.
Nonreact***	Tendency not to react to one's experience.	I watch my feelings without getting lost in them.
PHLMS		
Awareness*	Tendency to be highly aware of one's internal and external experiences.	When I am startled, I notice what is going on inside my body.
Acceptance**	Tendency to accept and not to judge internal and external experiences.	I try to put my problems out of mind.
TMS		
Curiosity*	Stance of wanting to learn more about one's experiences.	I am curious about each of my thoughts and feelings as they occur.

Decenter***	Tendency to relate to one's thoughts or feelings in a wider field of Awareness rather than being overly absorbed in one's internal experiences.	I experience myself as separate from my changing thoughts and feelings.

Note. Conceptually similar facets are denoted by the number of asterisks. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); PHLMS = Philadelphia Mindfulness Scale (Cardaciotto et al., 2008); TMS = Toronto Mindfulness Scale (Davis et al., 2009).

3.1.2. Convergent validity of measures and linkages to the Five-Factor Model

Associations among mindfulness scales have been examined in only a few studies, which have examined scale interrelations of only some of the measures, often with the aim of validating a particular scale. Baer et al. (2006) reported intercorrelations of five mindfulness scales, all of which were within a moderate range of .31 to .67. As can be expected, the two lowest correlations were reported for a relatively narrow measure, the MAAS (Brown & Ryan, 2003), which focuses on mindfulness attention and awareness. Intercorrelations of two particular measures with several others were also within a moderate range (Chadwick et al., 2008; Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2006). The distinct LMS showed weak-to-moderate correlations with two other scales (r = .27 to .37; Pirson et al., 2012). Generally, between-scale correlations have mostly been within a moderate range, which speaks to the differences in how the construct is operationalised across these unidimensional measures. It is unclear whether multiple dimensions explain the observed intercorrelations and, therefore, the shared variance among the scales.

Research into associations between mindfulness and the Big Five personality traits was reviewed in a meta-analysis of 32 samples by Giluk (2009). The focus of that study was exclusively on the mainstream conceptualisation of mindfulness, integrating the results from all relevant measures. Of the Big Five, Neuroticism was identified as the strongest correlate of mindfulness (r = -.58), followed by Conscientiousness (r = .44). Agreeableness also had an average correlation of moderate strength (r = .30), whereas Extraversion and Openness both correlated weakly with mindfulness (r = .10 and .07, respectively). However, the methodology of that review had several limitations.

One limitation is that the meta-analysis included data from studies that did not report all of the correlations between mindfulness and the Big Five. This practice may have biased the results, with statistical significance leading to the publication of only some of the Big Five's associations with mindfulness, thus inflating average intercorrelations. Another limitation was the inclusion of a bidimensional measure comprised of two orthogonal subscales (Cardaciotto et al., 2008). Weakly related factors, let alone unrelated ones, most likely represent multiple dimensions, and using them to measure a single construct has been described as indefensible (Smith et al., 2009). A third possible limitation was the inclusion of facet or subscale correlations with the Big Five, even though composite correlations of multiple facets and subscales with each personality dimension were calculated, presumably to address this problem. Since facet and especially subscale correlations with the Big Five are likely to vary (between each other and compared to global mindfulness), their inclusion may have led to inaccurate results in regards to global mindfulness. For example, not all mindfulness scales have facets or subscales, and it was not stated whether these correlations, where examined, were always reported for all factors or subscales. Moreover, facets or subscales are more likely than a global mindfulness score to comprise specific variance unrelated to the construct.

Correlations of the distinct LMS with the Five-Factor Model were reported in two studies. One of these studies only reported coefficients for Openness and Neuroticism (r = .73 and -.27, respectively; Pirson et al., 2012). In the other study, the measure's correlations with the Five-Factor Model factors were .50 with Openness, -.21 for Neuroticism, .35 for Extraversion, .23 for Conscientiousness, and .20 for Agreeableness (Bodner & Langer, 2001). This unique pattern of associations with the Big Five, revealing Openness as the strongest correlate, further speaks to the distinctiveness of the measure and the underlying construct. However, more evidence for the measure's linkages to the FFM, in direct comparison to measures aligned to the original conceptualisation of mindfulness, is needed.

In sum, several factors suggest that the relationship between mindfulness and the Five-Factor Model currently portrayed in the literature may not be fully accurate. First, differences in the construct validity between measures may distort our understanding of the true relationships. Second, very few studies, if any, have examined the relative "contributions" of relevant higher-order factors, such as the Big Five, to mindfulness. The relative contributions may well differ from the picture created by zero-order correlations, given that the Big Five are not perfectly orthogonal in a statistical sense (e.g., Van der Linden, Tsaousis, & Petrides, 2012). Last, the file-drawer phenomenon may have influenced the pattern of results reported in Giluk's (2009) meta-analysis, with non-significant relations (including those of facets and subscales) being underreported.

3.1.3. Facet and subscale correlations between measures

The PHLMS subscales were found to correlate weakly to moderately with the MAAS (Cardaciotto et al., 2008). Specifically, correlations with the MAAS were .21 in a non-clinical sample and .40 in a general psychiatric sample for PHLMS Awareness, whereas PHLMS Acceptance correlations with the MAAS were .32 in the non-clinical sample and .17 (non-significant) in the clinical sample. Correlations with facet scores of a multi-faceted measure, the KIMS, were mostly in line with conceptual similarities: PHLMS Awareness correlated strongly with the KIMS Observe facet (r = .83) and PHLMS Acceptance had the strongest correlation with the KIMS Accept without Judgment facet (r = .79).

Both subscales of the TMS (trait version) were associated with the unidimensional facet or item scales, ranging from .22 to .48 for TMS Curiosity and from .47 to .74 for TMS Decenter, as well as with the FFMQ and KIMS facets (Davis et al., 2009). TMS Curiosity correlated most highly with the FFMQ (r = .51) and KIMS (r = .54) Observe facet, as well as moderately with the FFMQ Nonreact facet (r = .32). Correlations of TMS Curiosity with the remaining facets were modest in strength (r = .10 to .21). TMS Decenter was most highly related to the FFMQ Nonreact facet (r = .74) and moderately to the other KIMS and FFMQ facets (Observe, Act with Awareness, and Accept without Judgment; r = .37 to .51), with the exception of the Describe facet ($r \leq .22$). Again, these values support the conceptual similarities between the facets and subscales of these measures.

Overall, there has been little empirical effort to systematically examine facet (or subscale) interrelationships and similarities among the key independently developed measures (FFMQ, PHLMS, and TMS) and to establish if all facets and subscales represent elements of the mainstream conceptualisation of mindfulness measured by the bulk of scales. A related specific concern is whether the FFMQ Observe facet, and possibly the conceptually and empirically related PHLMS Awareness and TMS Curiosity subscales, represent valid elements of this construct.

3.2. Study 1

The present study aimed to examine the homogeneity of existing mindfulness scales and establish whether a single dimension accounts for their shared variance. Two different samples completed all relevant trait measures that yield a global mindfulness score. A related aim was to investigate the linkages of conceptually and dimensionally distinct mindfulness scales to the Five-Factor Model in one of the samples, addressing some of the limitations of previous research. This aim served to solidify understanding of the level of similarity between existing scales and further elucidate any differences that may exist between underlying dimensions. In contrast to Giluk's (2009) meta-analysis, only global mindfulness scores were used, which implied the exclusion of the two multidimensional measures (PHLMS and TMS). The unique contributions of the Big Five to mindfulness were examined through the use of multiple regression analysis.

3.2.1. Method

3.2.1.1. Participants and procedure

Sample 1 (N = 397, 76.0% female) was recruited via the institutional subject pool of a major British university over approximately two years (February 2012–March 2014). The mean age was 21.9 years (SD = 5.0), ranging from 18.0 to 57.2 years. Predominantly comprising participants of White – UK heritage or other (53.1%), the sample also included participants from East Asian (29.6%) and South Asian (Indian, Pakistani, and Bangladeshi [8.3%]) backgrounds, as well as from multi-ethnic or other backgrounds (8.9%). The samples consisted mainly of undergraduate and Master’s students from various disciplines, predominantly from psychology and linguistics, but also included other individuals affiliated with the same institution. All participants were entered into a prize draw for gift cards and most received course credit for their participation.

Sample 2 (N = 176, 79.5% female) was recruited online using a twofold recruitment procedure in order to obtain a more heterogeneous sample with respect to mindfulness. First, a recruitment notice was posted on participant recruitment platforms for psychological research (e.g., http://www.onlinepsychresearch.co.uk/). Second, two promoters of mindfulness kindly agreed to post a recruitment notice on their social media pages. The average age of this sample (*M* = 36.37 years, *SD* = 14.4) was higher than that of Sample 1 and ranged from 15.7 to 76.2 years. Sample 2 was more homogeneous in terms of participant ethnic backgrounds, which were as follows: 84.1% Caucasian, 2.8% East Asian, 1.7% South Asian, 4.5% Black, and 6.8% other/mixed. A price draw of gift cards was offered to participants as a token of appreciation.

Participants of both samples provided demographic information and completed the mindfulness measures described in the next section via an anonymous electronic survey system. The Sample 1 participants additionally completed the Big Five measure described below. To balance the effects of any extraneous factors, such as testing fatigue, the scales were administered in randomised order, and the order of items within each scale was also randomised across participants. Upon submitting their responses on each scale, participants were automatically notified of any missing responses and given the opportunity to add them.

3.2.1.2. Measures

All instruments were based on self-report, multiple-point response scales, and showed good levels of internal reliability. Internal consistencies for the mindfulness scales are shown in Table 3.2, whereas those for the Big Five are included within the relevant description below.

Mindfulness

Five Facet Mindfulness Questionnaire (FFMQ; Baer et al., 2006). The FFMQ was developed as a comprehensive measure of the construct, by factor-analysing all of the scales below, except for the LMS (Bodner & Langer, 2001), which is based on a divergent conceptualisation of mindfulness. This procedure resulted in 39 items distributed across five facets (Observe, Describe, Act with Awareness, Accept without

Judgment, and Nonreact). The FFMQ items are rated on a 5-point Likert scale, ranging from 1 (*never or very rarely true*) to 5 (*very often or always true*).

Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004). The KIMS, which also comprises 39 items, is divided into four facets: Observe, Describe, Act with Awareness, and Accept without Judgment. All four facets and 24 of the 39 items are now contained within the FFMQ. The KIMS is based on the same 5-point response scale as the FFMQ.

Cognitive and Affective Mindfulness Scale – Revised (CAMS–R; Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2006). The CAMS–R global score is also based on four facets (attention, present focus, awareness, and acceptance), each represented by three items (12 in total). However, the use of facet scores is not encouraged, given the small number of items (three) per facet. The items are rated on a 4-point Likert scale from 1 (*Rarely/Not at all*) to 4 (*Almost Always*).

Southampton Mindfulness Questionnaire (SMQ; Chadwick et al., 2008). The SMQ consists of 16 items, representing four aspects of mindfulness: mindful observation, letting go of reacting, opening awareness to difficult experience, and acceptance. Similar to the CAMS–R, the developers have advised against computing facet scores, which had low alphas in the present samples. The response scale of the SMQ ranges from 0 (*Disagree Totally*) to 6 (*Agree Totally*).

Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003). The MAAS focuses exclusively on attentional aspects of mindfulness, whereas other scales also incorporate emotional aspects. Fifteen items are responded to on a 6-point Likert scale, ranging from 1 (*Almost Always*) to 6 (*Almost Never*).

Freiburg Mindfulness Inventory (FMI; Walach, Buchheld, Buttenmüller, Kleinknecht, & Schmidt, 2006). The FMI measures mindfulness through 14 items, based on a response scale of 1 (*rarely*) to 4 (*almost always*). The items represent basic aspects of mindfulness: attention to present moment (presence) and non-judgemental attitude (acceptance; Kohls, Sauer, & Walach, 2009). Although two highly interrelated factors have been derived from items of this measure (Kohls et al., 2009), factor scores or subscales have not been used in research.

Langer Mindfulness Scale (LMS; Bodner & Langer, 2001; Pirson et al., 2012). A revised 14-item version of the LMS (Pirson et al., 2012), which is grounded in a different conceptualisation of mindfulness, was used in this study. The items are distributed across three areas (Novelty seeking, engagement, and novelty producing) and responded to on a Likert scale, ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*).

Personality

The Big Five Inventory (John & Srivastava, 1999) was selected as a measure of the FFM. Forty-four brief descriptive items are responded to on a 5-point scale, ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Internal reliabilities were .85 for Neuroticism, .85 for Extraversion, .81 for Openness, .71 for Agreeableness, and .79 for Conscientiousness.

3.2.1.3. Statistical analyses

After computing intercorrelations among mindfulness scales, it was examined if more than a single dimension underlies the shared variance of the mindfulness scales. Excluded from these analyses was the FFMQ, as it derives from the other five scales based on the mainstream conceptualisation of the construct. Including the FFMQ in these analyses would duplicate the content of these five measures and bias the results against the LMS. The rest of the scales, including the LMS, were submitted to a Principal Component Analysis.

Bivariate correlations between mindfulness scales and the Big Five as well as average correlations of each Big Five trait with these scales were examined. The LMS was excluded from the average correlations, due to its distinct conceptualisation. To assess the unique contributions of the Big Five to mindfulness and the amount of overlap between the Five-Factor Model and mindfulness, regression analyses were conducted.

3.2.2. Results

3.2.2.1. Intercorrelations among mindfulness scales

Intercorrelations among the mindfulness scales are shown in Table 3.2. These were consistent between the two samples in that for all scales, except the LMS, coefficients exceeded .30. The only correlation below this level was between the SMQ and MAAS in Sample 1 (r = .24). Still, the magnitude of the correlations varied widely: .25 to .90 in Sample 1 and .36 to .95 in Sample 2. In contrast, correlations between the LMS and the other scales were generally weaker, reflecting the developers' distinct conceptualisation of the construct. Specifically, the LMS showed weak average correlations with the other scales in both Sample 1 (r = .19, range = .00 to .33) and Sample 2 (r = .27, range = .16 to .39).

Table 3.2. *Study 1: Internal Consistencies and Intercorrelations among Mindfulness Scales*

	FFMQ	KIMS	CAMS–R	SMQ	MAAS	FMI	LMS
			Sample 1				
FFMQ	(.84)						
KIMS	.90***	(.80)					
CAMS–R	.67***	.60***	(.74)				
SMQ	.50***	.34***	.52***	(.80)			
MAAS	.52***	.46***	.44***	.25***	(.86)		
FMI	.59***	.49***	.60***	.49***	.34***	(.83)	
LMS	.33***	.33***	.14**	.00	.11*	.21***	(.82)
			Sample 2				
FFMQ	(.92)						
KIMS	.95***	(.89)					
CAMS–R	.77***	.72***	(.83)				
SMQ	.72***	.66***	.64***	(.87)			
MAAS	.60***	.59***	.49***	.36***	(.88)		
FMI	.70***	.61***	.75***	.69***	.48***	(.89)	
LMS	.36***	.39***	.25***	.24**	.16	.20*	(.86)

Note. N = 397 for Sample 1; N = 176 for Sample 2, but only 120 participants completed the MAAS and FMI in Sample 2. Coefficient alphas are reported in parentheses along the diagonal for each sample. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); CAMS–R = Cognitive and Affective Mindfulness Scale – Revised (Feldman et al., 2006); SMQ = Southampton Mindfulness Questionnaire (Chadwick et al., 2008); MAAS = Mindful Attention Awareness Scale (Brown & Ryan, 2003); FMI = Freiburg Mindfulness Inventory (Walach et al., 2006); LMS = Langer Mindfulness Scale (Bodner & Langer, 2001; Pirson et al., 2012).
*$p < .05$. **$p < .01$. ***$p < .001$.

With all six scales shown in Table 3.3 included in the Principal Component Analysis, two components emerged in Sample 1 and one component in Sample 2. Due to a high loading of the LMS and negligible loadings from the other scales, the LMS was mainly accountable for the second component in Sample 1 (two of the other scales loaded negatively on this component). Additionally, the LMS had relatively weak loadings on the first component in both samples (λ = .34 and .32), whereas the other scales showed relatively high loadings (λ = .63 to .86). These results and the distinct conceptualisation of mindfulness underlying the LMS led to a repetition of the analysis without the LMS. The results of the reanalysis are shown in parentheses in Table 3.3. Without the LMS, a single principal component accounted for the shared variance among the scales in both samples (56.7% and 67.4%), in each case explaining close to 10 percent more variance than the analyses containing the LMS.

Table 3.3. *Study 1: Principal Component Analyses of Mindfulness Scales*

Sample	Scale	Factor loading	Communality	% of variance
1	KIMS	.80 (.78)	.69 (.61)	48.40 (56.68)
	CAMS–R	.85 (.86)	.74 (.73)	
	SMQ	.66 (.68)	.64 (.67)	
	MAAS	.63 (.64)	.40 (.71)	
	FMI	.79 (.79)	.62 (.62)	
	LMS	.32	.86	
2	KIMS	.86 (.85)	.74 (.73)	57.60 (67.42)
	CAMS–R	.86 (.87)	.75 (.76)	
	SMQ	.79 (.80)	.63 (.64)	
	MAAS	.69 (.69)	.47 (.48)	
	FMI	.86 (.87)	.75 (.76)	
	LMS	.34	.12	

Note. N = 397 for Sample 1 and 120 for Sample 2. Results shown in parentheses derive from analyses excluding the LMS, which loaded highly on a second component in Sample 1 (λ = .87) and relatively weakly on the first component in both samples. KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); CAMS–R = Cognitive and Affective Mindfulness Scale – Revised (Feldman et al., 2006); SMQ = Southampton Mindfulness Questionnaire (Chadwick et al., 2008); MAAS = Mindful Attention Awareness Scale (Brown & Ryan, 2003); FMI = Freiburg Mindfulness Inventory (Walach et al., 2006); LMS = Langer Mindfulness Scale (Bodner & Langer, 2001; Pirson et al., 2012).

3.2.2.2. Mindfulness and the Big Five

Bivariate correlations between mindfulness scales and the Big Five are shown in Table 3.4. Extraversion and Conscientiousness correlated with all of the mindfulness scales. Neuroticism correlated with all of the scales based on the mainstream conceptualisation, but not with the LMS. Agreeableness correlated with all scales except for the SMQ. Openness was the least reliable correlate of the mindfulness scales based on the mainstream conceptualisation; it correlated with the FFMQ, KIMS, FMI, but not with the CAMS–R, SMQ, and MAAS. In contrast, it was the strongest personality correlate of the LMS. All significant correlations were in an expected direction. Neuroticism was the only Big Five dimension showing moderately strong correlations with all mindfulness scales based on the mainstream conceptualisation (r = -.32 to -.58). The other four dimensions showed a mix of weak-to-moderate correlations (r = .12 to .42). The LMS' correlation with Openness was the strongest in the matrix (r = .67). However, its other significant correlations with personality dimensions were all relatively weak (r = .15 to .24).

Table 3.4. *Study 1: Bivariate Correlations between Mindfulness Scales and the Big Five in Sample 1*

	FFMQ	KIMS	CAMS–R	SMQ	MAAS	FMI	LMS
Neuroticism	-.47***	-.32***	-.52***	-.58***	-.35***	-.55***	-.08
Extraversion	.34***	.32***	.15**	.16**	.14**	.24***	.24***
Openness	.31***	.35***	.05	-.01	.02	.21***	.67***
Agreeableness	.27***	.26***	.21***	.08	.31***	.22***	.15**
Conscientiousness	.37***	.36***	.42***	.12*	.31***	.16**	.19***

Note. N = 358. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); CAMS–R = Cognitive and Affective Mindfulness Scale – Revised (Feldman et al., 2006); SMQ = Southampton Mindfulness Questionnaire (Chadwick et al., 2008); MAAS = Mindful Attention Awareness Scale (Brown & Ryan, 2003); FMI = Freiburg Mindfulness Inventory (Walach et al., 2006); LMS = Langer Mindfulness Scale (Bodner & Langer, 2001; Pirson et al., 2012).
*p < .05. **p < .01. ***p < .001.

Average correlations of the Big Five with the mindfulness scales, excluding the LMS, were as follows: -.46 for Neuroticism, .22 for Extraversion, .15 for Openness, .22 for Agreeableness, and .29 for Conscientiousness.

The considerable variability in magnitude of associations among mindfulness scales reflects previous findings and suggests that the scales vary in the degree to which they measure the construct. Consequently, linkages of mindfulness to the Five-Factor Model were not separately examined for all scales, since differences in the breadth of these measures could lead to divergent patterns of associations and uncertainty about the relationships between mindfulness and the FFM. Since all scales loaded on a single component, a composite of the KIMS, CAMS–R, SMQ, MAAS, and FMI was derived from the Principal Component Analysis described above, excluding the LMS. The FFMQ was examined separately as a way of cross-validation; it derives from these five scales and showed good convergence with their composite at .85 in Sample 1 and .90 in Sample 2. The LMS' linkages to the Big Five were also examined in a separate analysis due to the distinct conceptualisation of mindfulness underlying this scale.

The regression analysis results are summarised in Table 3.5. Beta weights for the Big Five had the same order of magnitude between the FFMQ and the multi-scale composite. Specifically, the order of predictors in terms of strength was Neuroticism, Conscientiousness, Openness, Extraversion, and Agreeableness. Extraversion was a significant predictor of the FFMQ only and Agreeableness did not show a significant effect on either variable. The remaining personality dimensions reached significance in both analyses. Overall, personality explained 43% and 51% of the mindfulness variance in the FFMQ and multi-scale composite scores, both of which represent the mainstream conceptualisation of mindfulness.

While personality explained a similar amount of variance in the LMS (49%), which seems to mainly assess a different construct, a very different pattern of predictive effects was observed for this measure. In this case, Openness was by far the strongest predictor, followed by Conscientiousness and Extraversion. The beta weights for Neuroticism and Agreeableness were not significant.

Table 3.5. *Study 1: Regressions of the FFMQ, Multi-Scale Composite, and LMS on the Big Five in Sample 1*

	FFMQ		MSC		LMS	
	$F(5,352) = 54.04$**		$F(5,352) = 72.11$**		$F(5,352) = 68.22$**	
Predictor	β	R^2	β	R^2	β	R^2
Neuroticism	-.38**	.43	-.56**	.51	-.04	.49
Extraversion	.17**		.07		.11*	
Openness	.28**		.15**		.65**	
Agreeableness	.05		.05		.02	
Conscientiousness	.28**		.27**		.16**	

Note. $N = 358$. Regression coefficients (β) represent standardised beta weights. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); MSC = multi-scale composite; LMS = Langer Mindfulness Scale (Bodner & Langer, 2001; Pirson et al., 2012).
*$p < .01$. **$p < .001$.

3.2.3. Discussion

The present study aimed to clarify issues surrounding the conceptualisation and measurement of dispositional mindfulness, particularly the similarity of the extant measures. The first issue concerned the convergent validity of the measures. Although correlations among measures aligned to the mainstream conceptualisation were generally within a moderate-to-strong range, there were considerable discrepancies. This pattern has been observed in previous studies (Baer et al., 2006; Chadwick et al., 2008; Feldman et al., 2006) and suggests that some measures represent the construct partially, while others represent it more comprehensively. Intercorrelations involving the LMS were noticeably lower than those of the other scales, as could be expected given its distinct conceptualisation of mindfulness and previous findings (Pirson et al., 2012). These results indicate that the LMS shares the least amount of variance with the other measures.

The second issue concerned whether a single dimension can account for the shared variance between mindfulness scales. The results from both samples showed that the shared variance of the scales based on the mainstream conceptualisation of mindfulness is explained by a single dimension, which presumably represents the target construct. In contrast, and consistent with the bivariate correlations across the two samples, the LMS loaded relatively weakly on this factor and even produced a second component in Sample 1, on which it loaded highly. These results strongly suggest that the two conceptualisations of mindfulness represent distinct constructs.

The third issue concerned the pattern of relationships between the various measures of mindfulness and the Big Five personality dimensions. Previous research has been mostly restricted to the mainstream conceptualisation, with a heterogeneous set of scales imposing some limitations to the interpretability of findings. Results conformed to Giluk's (2009) meta-analysis in that Neuroticism showed the strongest, and Conscientiousness the second strongest, relationship with the multi-scale composite and FFMQ total scores. On the other hand, Giluk’s meta-analysis revealed Extraversion as the weakest correlate, whereas the weakest average correlate in the present sample was Openness; Extraversion showed the same magnitude of association as Agreeableness, which was the third strongest correlate in Giluk’s meta-analysis. These differences may have several explanations. First and foremost, the present results involving the Big Five are based on a single sample and on a single measure of the Big Five traits, whereas Giluk integrated the results of multiple samples spanning various Big Five measures. On the other hand, as mentioned in the introduction, Giluk’s meta-analysis had certain limitations, including possible file-drawer effects and the inclusion of a mindfulness scale comprised of orthogonal factors.

An advantage of the present investigation is that it examined the unique contributions of the Big Five to dispositional mindfulness. Since the five unidimensional scales based on the mainstream conceptualisation of mindfulness loaded on a single component, a multi-scale composite (rather than each constituent scale) was used in the present study to examine the linkages of the underlying dimension to the Big Five. The strategic benefit of this approach was that this composite should yield a more comprehensive representation of the construct and reveal its linkages to the Five-Factor Model more accurately than individual measures. In addition, the FFMQ was examined separately, because it was empirically derived from these scales (Baer et al., 2006) and, thus, useful for cross-validation purposes.

When regressing the two very similar variables representative of the mainstream conceptualisation (the FFMQ global score and the composite derived from the other unidimensional scales) on the Big Five, a slightly different picture emerged compared to the zero-order correlations. While Neuroticism and Conscientiousness remained the strongest predictors, Openness, which showed the weakest average correlation, became the third strongest predictor of the multi-scale composite and, together with Conscientiousness, the second strongest predictor of the FFMQ global score. Surprisingly, Agreeableness had no unique predictive effects on either variable. Extraversion predicted the FFMQ, but not the multi-scale composite. The two mindfulness scores shared about half their variance with the FFM.

The LMS's pattern of associations with the Five-Factor Model was very different from that observed for the FFMQ and multi-scale composite scores. Neuroticism was the weakest and sole non-significant correlate, despite previous reports of small, but significant, correlations with the LMS (Bodner & Langer, 2001; Pirson et al., 2012). Openness (the weakest average correlate of the other variables) was by far the strongest correlate of the LMS. The strong association with Openness is not surprising, given the nature of the model and its facets (novelty producing, novelty seeking, and engagement), which reflect this basic personality dimension. Also, similar associations with Openness were previously reported in Pirson et al. (2012; r = .73) and Bodner and Langer (2001; r = 50). The remaining Big Five traits had significant, but relatively weak, correlations with the LMS, again of similar magnitude as correlations reported previously (Pirson et al., 2012). Regression analysis suggested a similar conclusion, except that Agreeableness did not predict unique LMS variance with the other four personality dimensions in the regression equation. These results suggest that Agreeableness is not uniquely related to mindfulness.

3.3. Study 2

The purpose of the present study was to examine the similarities, validity, and dimensionality of the FFMQ facets and PHLMS and TMS subscales. The analyses conducted to address these questions included: (1) intercorrelations among the FFMQ facets and PHLMS and TMS subscales; (2) bivariate correlations of PHLMS and TMS subscales with mindfulness, a composite variable derived from the unidimensional facet or item scales (other than the FFMQ and LMS); and (3) joint Exploratory and Confirmatory Factor Analysis of these three measures. Prior to the third part of the analyses (joint Confirmatory Factor Analysis), four- and five-factor hierarchical models were tested for the FFMQ, given the inconsistencies previously found with the Observe facet (the PHLMS and TMS each comprise largely distinct subscales that do not combine into a global mindfulness factor).

3.3.1. Method

3.3.1.1. Participants

The two samples used in Study 1 were used here, but minus a few cases with missing items on relevant facets or subscales. The effective sample sizes were 396 (76.6% female) for Sample 1 and 172 (79.7% female) for Sample 2.

3.3.1.2. Measures and procedure

Of the mindfulness scales used and described in Study 1, the following were used: FFMQ (Baer et al., 2006), KIMS (Baer, Smith, & Allen, 2004), CAMS–R (Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2006), SMQ (Chadwick et al., 2008), MAAS (Brown & Ryan, 2003), and FMI (Walach, Buchheld, Buttenmüller, Kleinknecht, & Schmidt, 2006). Data for the MAAS and FMI were only available for

115 of the Sample 2 participants. Furthermore, participants completed the PHLMS (Cardaciotto et al., 2008), which comprises two orthogonal subscales (Awareness and Acceptance). The PHLMS items are rated on a 5-point response scale, ranging from 1 (*Never*) to 5 (*Very Often*). Participants also completed the second bidimensional measure, the TMS (trait version; Davis, Lau, & Cairns, 2009), which comprises the subscales of Curiosity and Decenter. The TMS items are responded to on a 4-point Likert scale, ranging from 1 (*Not at all*) to 5 (*Very much*). Properties of these measures, including number of items and internal consistencies in the two study samples, are shown Table 3.6. As indicated, the levels of internal reliability range from acceptable to strong in both samples and across measures. All scales were administered as described in Study 1.

Table 3.6. *Study 2: Descriptive Statistics and Properties of Mindfulness Scales*

		Sample 1 (N = 395)					Sample 2 (N = 172)				
Scales	No. of items	α	*M*	*SD*	Skewness	Kurtosis	α	*M*	*SD*	Skewness	Kurtosis
FFMQ Observe	8	.80	3.21	0.69	0.13	-0.13	.82	3.48	0.66	-0.31	0.51
FFMQ Describe	8	.88	3.26	0.74	-0.01	-0.42	.93	3.39	0.90	-0.27	-0.35
FFMQ Act with Awareness	8	.88	3.10	0.73	-0.08	0.16	.91	2.95	0.77	0.10	-0.24
FFMQ Accept w/o Judgment	8	.91	3.00	0.85	-0.11	-0.31	.94	2.87	0.99	0.22	-0.60
FFMQ Nonreact	7	.83	2.86	0.68	0.22	0.10	.88	2.70	0.72	0.11	-0.14
PHLMS Awareness	10	.78	3.51	0.55	-0.06	0.06	.85	3.69	0.62	-0.44	0.47
PHLMS Acceptance	10	.82	2.71	0.64	-0.08	-0.16	.88	2.71	0.76	0.05	-0.59
TMS Curiosity	6	.86	2.50	0.80	-0.32	-0.12	.88	2.51	0.83	-0.41	-0.20
TMS Decenter	7	.73	1.87	0.67	-0.01	-0.07	.77	1.71	0.73	0.00	-0.37
KIMS	39	.81	3.06	0.35	0.25	0.91	.89	3.13	0.46	0.01	0.76
CAMS–R	12	.75	2.55	0.43	-0.08	0.09	.83	2.46	0.50	0.21	-0.32
SMQ	16	.80	3.25	0.74	-0.36	0.06	.87	3.02	0.91	-0.11	0.05
MAAS	15	.86	3.67	0.72	0.11	0.19	.88	3.60	0.80	-0.20	0.06
FMI	14	.83	2.57	0.47	0.01	0.32	.89	2.40	0.59	0.16	-0.39

Note. Of the Sample 2 participants, only 115 completed the MAAS and FMI. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); PHLMS = Philadelphia Mindfulness Scale (Cardaciotto et al., 2008); TMS = Toronto Mindfulness Scale (Davis et al., 2009); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); CAMS–R = Cognitive and Affective Mindfulness Scale – Revised (Feldman et al., 2006); SMQ = Southampton Mindfulness Questionnaire (Chadwick et al., 2008); MAAS = Mindful Attention Awareness Scale (Cardaciotto et al., 2008); FMI = Freiburg Mindfulness Inventory (Walach et al., 2006).

3.3.1.3. Statistical analyses

Intercorrelations among the FFMQ facets and PHLMS and TMS subscales were examined to assess their convergent/divergent validities and similarities. Using Principal Component Analysis, a comprehensive mindfulness component was extracted from the KIMS, CAMS–R, SMQ, MAAS, and FMI (cf. Section 3.2 - Study 1). The FFMQ, which derives from these five measures, was not included in this composite. Bivariate correlations of the PHLMS and TMS subscales with this mindfulness component were computed to examine the extent to which these subscales map onto the mainstream conceptualisation of mindfulness implicit in most of the extant measures.

Confirmatory Factor Analysis with maximum-likelihood estimates was used to (a) test the five- and four-factor hierarchical models of the FFMQ (the respective sample sizes can be considered sufficient for this purpose, given numbers of parameters; Bentler & Chou, 1987), and (b) examine which of the PHLMS and TMS subscales load on the same factor as the FFMQ facets and, thus, represent the same construct. These analyses were conducted on item parcels, as executed in the construction of the FFMQ (Baer et al., 2006) and in later tests of its factor structure (e.g., Williams et al., 2014). Specifically, three parcels per facet were created, by assigning the items in the order in which they appear in the FFMQ across parcels (e.g., Describe Item 1 → Describe Parcel 1, Describe Item 2 → Describe Parcel 2, and so forth). Since justifications for the use of items parcels in this context were previously presented in Baer et al. (2006), they will not be reiterated here.

Objective (a) was examined in both samples in order to establish the robustness of the results and to maximise justification for the analytical design pertaining to Objective (b) in terms of variables included (excluded). Objective (b) was approached by a joint Exploratory Factor Analysis of item parcels in Sample 2 (the smaller sample), using principal axis factoring with oblique rotation (Promax method, delta = 4), followed by a joint Confirmatory Factor Analysis of the extracted model in Sample 1, using maximum-likelihood estimation. In line with contemporary thinking on adequate model fit (Byrne, 1994; Hooper, Coughlan, & Mullen, 2008; Iacobucci, 2010), the following indices were used (leniently): Goodness of Fit Index (GFI), Comparative Fit Index (CFI), Normed Fit Index (NFI), and Standardised Root Mean Square Residual (SRMR); the criteria for the various fit indices were: GFI $\geq$ .90, CFI $\geq$ .93, NFI $\geq$ .90, and SRMR $\leq$ .09.

3.3.2. Results

3.3.2.1. Correlations

Zero-order correlations for the FFMQ facets, PHLMS and TMS subscales, and mindfulness component are shown in Table 3.7. Strengths of significant associations among these scores appear to be generally stronger in Sample 2. Concerning the FFMQ, facets were generally non-significantly or weakly associated in Sample 1 and weakly to moderately in Sample 2. Of note, the Observe facet showed significant negative correlations with Act with Awareness and Accept without Judgment, whereas the same correlations were non-significant in Sample 2.

The PHLMS subscales were inversely related in both samples, although coefficients were small and only significant in Sample 1 at r = -.28. PHLMS Awareness correlated moderately with the FFMQ Observe and Describe facets in both samples. Its association with the other facets were non-significant or negative (r = -.11) in Sample 1 and weak or non-significant in Sample 2. In both samples, the strongest FFMQ correlate of PHLMS Acceptance was Accept without Judgment at r = .58 and .72. PHLMS Acceptance also correlated moderately with FFMQ Act with Awareness and weakly to moderately with FFMQ Nonreact in both samples. Its association with FFMQ Describe was non-significant in Sample 1 and weak in Sample 2. Moreover, the association between PHLMS Acceptance and FFMQ Observe was negative in Sample 1 and non-significant in Sample 2.

The two TMS subscales correlated moderately in both samples (r = .39 and .40). TMS Curiosity also correlated moderately with FFMQ Observe and weakly with the Describe and Nonreact facets in both samples. Its associations with the other two FFMQ facets were negative and significant in only one instance (a negative correlation with Accept without Judgment in Sample 1). TMS Curiosity correlated moderately with PHLMS Awareness in both samples and either negatively or non-significantly with PHLMS Acceptance. In contrast, TMS Decenter correlated most strongly with FFMQ Nonreact in both samples (r = .51 and .55), followed by FFMQ Observe (r = .33 and .29). FFMQ Describe and Act with Awareness were both unrelated to TMS Decenter, whereas Accept without Judgment correlated weakly in Sample 2. Correlations of TMS Decenter were weak with PHLMS Awareness and significant in Sample 2 for PHLMS Acceptance (r = .32).

Correlations of the PHLMS and TMS subscales with mindfulness were consistently significant and positive. Most correlations were within a moderate range, but TMS Curiosity and PHLMS Awareness were weaker correlates of this composite, although the association for PHLMS Awareness in Sample 2 was also of moderate degree.

Table 3.7. *Study 2: Intercorrelations among FFMQ Facets, PHLMS and TMS Subscales, and the Global Mindfulness Component*

	Mindfulness	1	2	3	4	5	6	7	8
				Sample 1 (*N* = 395)					
1. FFMQ Observe	—								
2. FFMQ Describe	—	.23***							
3. FFMQ Act with Awareness	—	-.11*	.08						
4. FFMQ Accept w/o Judgment	—	-.22***	.09	.44***					
5. FFMQ Nonreact	—	.23***	.07	.09	.13*				
6. PHLMS Awareness	.26***	.62***	.40***	.04	-.11*	.09			
7. PHLMS Acceptance	.41***	-.20***	.07	.39***	.58***	.16**	-.28***		
8. TMS Curiosity	.12*	.40***	.17***	-.04	-.21***	.10*	.42***	-.14**	
9. TMS Decenter	.38***	.33***	.06	.02	-.02	.51***	.24***	.00	.39***
				Sample 2 (*N* = 172)					
1. FFMQ Observe	—								
2. FFMQ Describe	—	.31***							
3. FFMQ Act with Awareness	—	.13	.30***						
4. FFMQ Accept w/o Judgment	—	.05	.21**	.56***					
5. FFMQ Nonreact	—	.36***	.17*	.27***	.45***				
6. PHLMS Awareness	.42***	.67***	.41***	.17*	.04	.25***			

7. PHLMS Acceptance	.63***	.06	.25***	.53***	.72***	.46***	-.04		
8. TMS Curiosity	.24*	.35***	.24**	-.03	-.03	.24**	.30***	.07	
9. TMS Decenter	.43***	.29***	.14	.12	.26***	.55***	.14	.34***	.40***

Note. Global mindfulness correlations in Sample 2 are based on the data of 115 participants, who completed all of the mindfulness measures in that sample. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); PHLMS = Philadelphia Mindfulness Scale (Cardaciotto et al., 2008); TMS = Toronto Mindfulness Scale (Davis et al., 2009).
*p < .05. **p < .01. ***p < .001.

3.3.2.2. Confirmatory Factor Analysis of FFMQ hierarchical models

The five-factor hierarchical model adequately fit the data in Sample 1, $\chi^2(85) = 216.99$, $p < .001$, CFI = .96, GFI = .93, NFI = .94, SRMR = .08, but not particularly well in Sample 2, $\chi^2(85) = 170.37$, $p < .001$, CFI = .95, GFI = .89, NFI = .91, SRMR = .10. However, FFMQ Observe loaded negatively on the latent mindfulness factor in Sample 1 ($\lambda = -.17$, $p = .21$), whereas in Sample 2, its loading on mindfulness was relatively small ($\lambda = .22$, $p = .02$). In contrast, the four-factor hierarchical model without the Observe facet fit the data very well in both Sample 1, $\chi^2(50) = 88.51$, $p = .001$, CFI = .99, GFI = .96, NFI = .97, SRMR = .04, and Sample 2, $\chi^2(50) = 61.28$, $p = .13$, CFI = .95, GFI = .99, NFI = .96, SRMR = .06.

Factor loadings for the better supported four-factor model are shown in Figure 3.1 and Figure *3.2*, pertaining to Samples 1 and 2, respectively. Although this model fit the data well in both samples, magnitudes of the standardised path coefficients were heterogeneous at the facet level. Specifically, facet loadings were low for Describe in both samples (especially in Sample 1, where it was not significant) and for Nonreact in Sample 1; they were relatively high for Act with Awareness and Accept without Judgment in both samples.

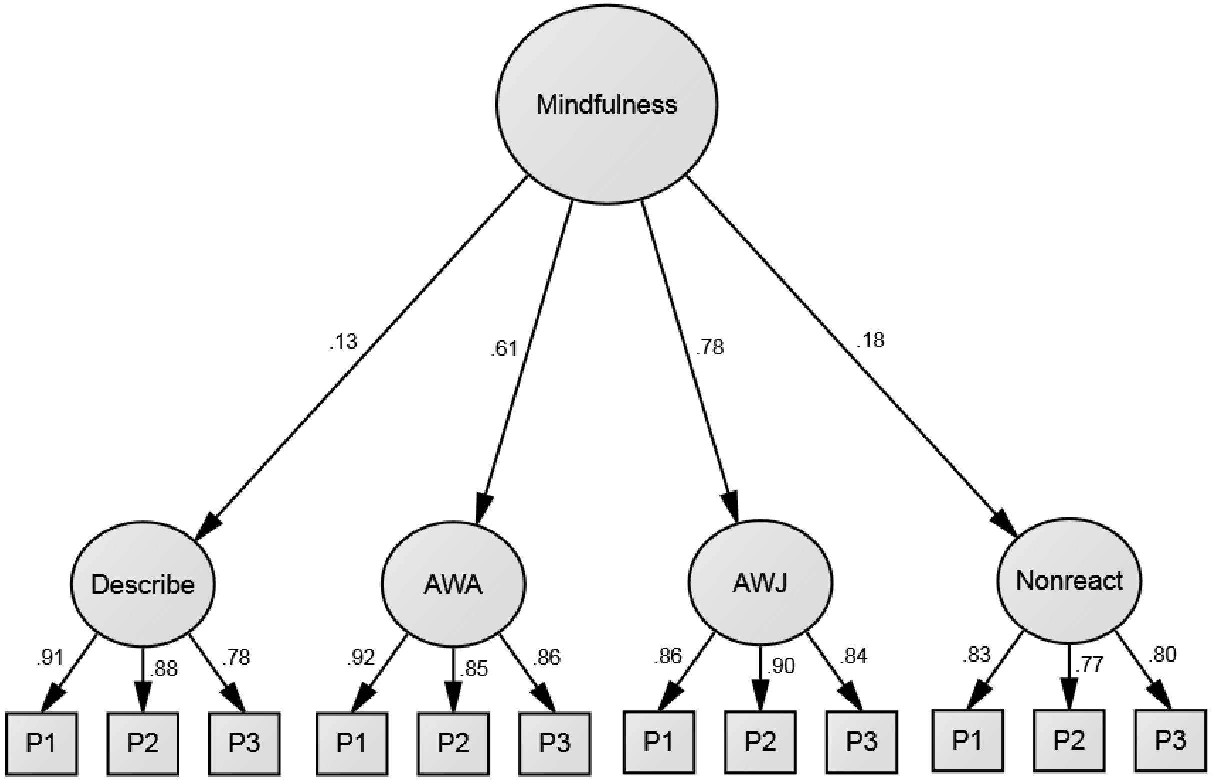


Figure 3.1. Study 2: Confirmatory Factor Analysis results for the four-factor hierarchical model of the Five Facet Mindfulness Questionnaire (Baer et al., 2006), omitting the Observe facet, in Sample 1 (N = 395). First-order latent variables represent the four facets and derive from item parcels (three per facet). Error terms are omitted for visual clarity. AWA = Act with Awareness; AWJ = Accept w/o Judgment; P1 to P3 = Parcels 1 to 3. All standardised coefficients are significant at the .05 level, with the exception of the path from Mindfulness to Describe, which did not reach significance (p = .09).

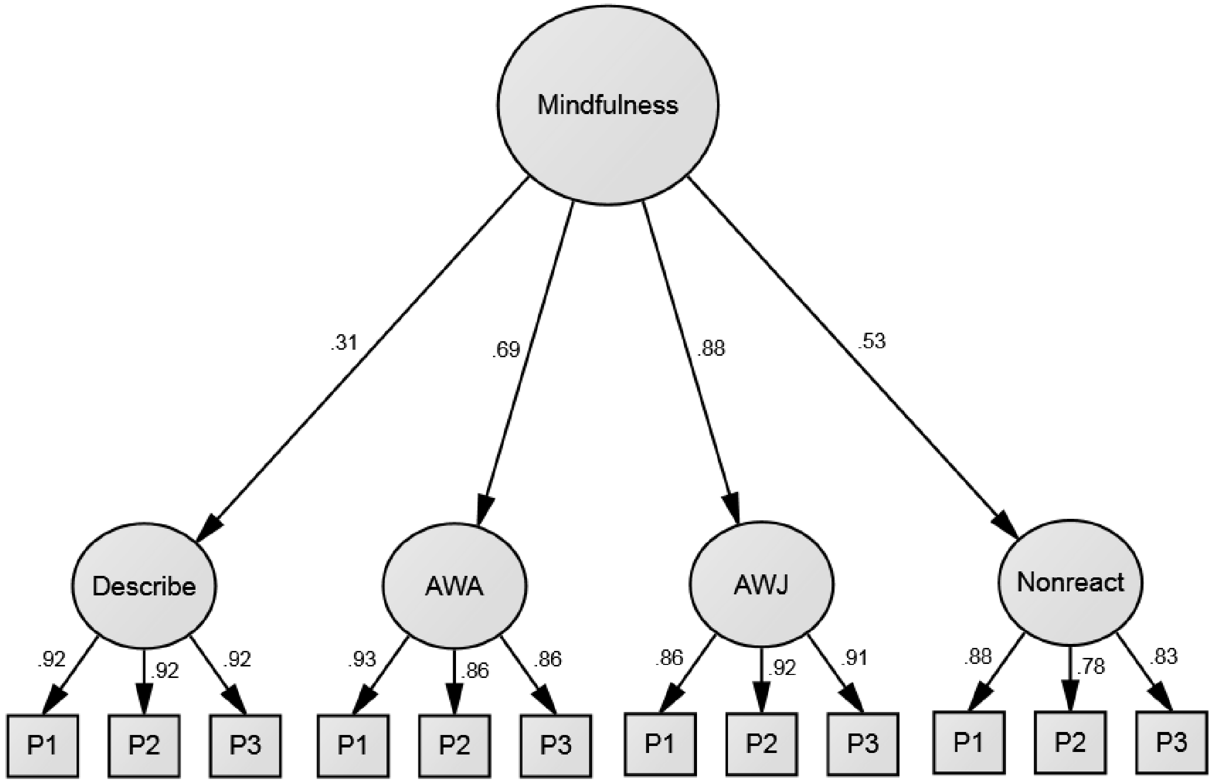


Figure 3.2. Study 2: Confirmatory Factor Analysis results for the four-factor hierarchical model of the Five Facet Mindfulness Questionnaire (Baer et al., 2006), omitting the Observe facet, in Sample 2 (*N* = 172). First-order latent variables represent the four facets and derive from item parcels (three per facet). Error terms are omitted for visual clarity. AWA = Act with Awareness; AWJ = Accept w/o Judgment; P1 to P3 = Parcels 1 to 3. All standardised coefficients are significant at the .01 level.

3.3.2.3. Exploratory Factor Analysis of FFMQ, PHLMS, and TMS

Given the preceding results, this analysis was conducted without the item parcels of FFMQ Observe. A scree plot did not clearly indicate a particular number of factors (see Appendix 1), while Kaiser's criterion (Eigenvalues > 1) supported a six-factor solution for the remaining FFMQ parcels combined with the PHLMS and TMS parcels. Indeed, six clean factors are apparent from the pattern matrix shown in Table 3.8, with any loadings below the conventional .30 cut-off suppressed; no cross-loadings above .30 appeared. Half the facets and subscales emerged as a distinct factor, characterised by loadings of all three respective item parcels: FFMQ Describe, FFMQ Act with Awareness, TMS Curiosity, and PHLMS Awareness. The two other factors were combinations of (a) FFMQ Accept without Judgment and PHLMS Acceptance and (b) FFMQ Nonreact and TMS Decenter; each factor was identified by the six respective parcels. The solution explained much of the variance in the parcels (76.95%).

On the other hand, a parallel analysis indicated an eight-factor solution. With the number of factors fixed to eight, every facet or subscale (eight in total) emerged as a distinct factor, characterised by loadings of the respective parcels. The output for this analysis is included in Appendix 1. For this reason, it made sense to examine two slightly different models via Confirmatory Factor Analysis in Sample 1; one comprised of six, and the other of eight, first-order factors between the item parcels and the second-order, global mindfulness factor. Of note, both models comprise the item parcels of all eight facets or subscales, but in the six-factor model, four of them emerged in pairs as two factors (see Table 3.8).

Table 3.8. Study 2: *Pattern Matrix for Promax Six-Factor Solution Extracted from FFMQ, TMS, and PHLMS Items Parcels Corresponding to Each Facet or Subscale and Factor Correlation Matrix in Sample 2*

	Factor loading					
Parcel	1	2	3	4	5	6
FFMQ AWJ P2	1.00					
FFMQ AWJ P3	.88					
FFMQ AWJ P1	.84					
PHLMS Acceptance P2	.70					
PHLMS Acceptance P1	.56					
PHLMS Acceptance P3	.52					
FFMQ Nonreact P1		.83				
FFMQ Nonreact P3		.81				
FFMQ Nonreact P2		.76				
TMS Decenter P3		.59				
TMS Decenter P2		.51				
TMS Decenter P1		.49				
FFMQ Describe P1			.90			
FFMQ Describe P2			.86			
FFMQ Describe P3			.83			
TMS Curiosity P3				.86		
TMS Curiosity P2				.85		
TMS Curiosity P1				.76		
FFMQ AWA P1					.91	
FFMQ AWA P3					.88	
FFMQ AWA P2					.83	
PHLMS Awareness P1						.72
PHLMS Awareness P3						.68
PHLMS Awareness P2						.68
Eigenvalue	5.67	4.44	2.86	1.76	1.57	1.18
% of variance	23.61	18.52	11.92	7.35	6.53	4.92

Factor correlations

Factor 1	—					
Factor 2	.29	—				
Factor 3	.21	.13	—			
Factor 4	-.13	.29	.20	—		
Factor 5	.55	.17	.19	-.03	—	
Factor 6	-.19	.10	.32	.30	.00	—

Note. *N* = 172. Factor loadings of < .30 are omitted from the table. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); PHLMS = Philadelphia Mindfulness Scale (Cardaciotto et al., 2008); TMS = Toronto Mindfulness Scale (Davis et al., 2009); AWJ = Accept w/o Judgment; AWA = Act with Awareness; P1 to P3 = Parcels 1 to 3.

3.3.2.4. Confirmatory Factor Analysis of FFMQ, PHLMS, and TMS

Initially, both models did not fit the data (six-factor model: $\chi^2[240] = 640.30$, $p < .001$, CFI = .93, GFI = .88, NFI = .88, SRMR = .11; eight-factor model: $\chi^2[244] = 759.57$, $p < .001$, CFI = .90, GFI = .86, NFI = .86, SRMR = .13) and contained two first-order factors that loaded negatively on the global mindfulness factor: PHLMS Awareness and TMS Curiosity. Upon removing these two latent variables and their constituent parcels from each model, the eight-factor model yielded satisfactory fit, $\chi^2(129) = 326.10$, $p < .001$, CFI = .95, GFI = .92, NFI = .92, SRMR = .09, whereas the six-factor model generally did not meet the specified criteria for model fit, $\chi^2(131) = 556.48$, $p < .001$, CFI = .89, GFI = .85, NFI = .86, SRMR = .07.

In the better fitting model, in which all remaining facets or subscales represent a distinct second-order factor, the loading for TMS Decenter (.04) was non-significant. Thus, Figure 3.3 shows a final model with TMS Decenter removed (and without any error covariances added), yielding very good fit, $\chi^2(85) = 161.43$, $p < .001$, CFI = .98, GFI = .95, NFI = .96, SRMR = .04. Loadings of the four FFMQ facets were largely unaffected by the additional PHLMS parcels in terms of magnitude with the PHLMS Acceptance factor included in the model (cf. Section 3.3.2.2.). Of note, the loading for FFMQ Describe became significant in this instance. The PHLMS Acceptance subscale loaded highly on mindfulness (.83), consistent with the high loading of the equivalent FFMQ facet (Accept without Judgment).

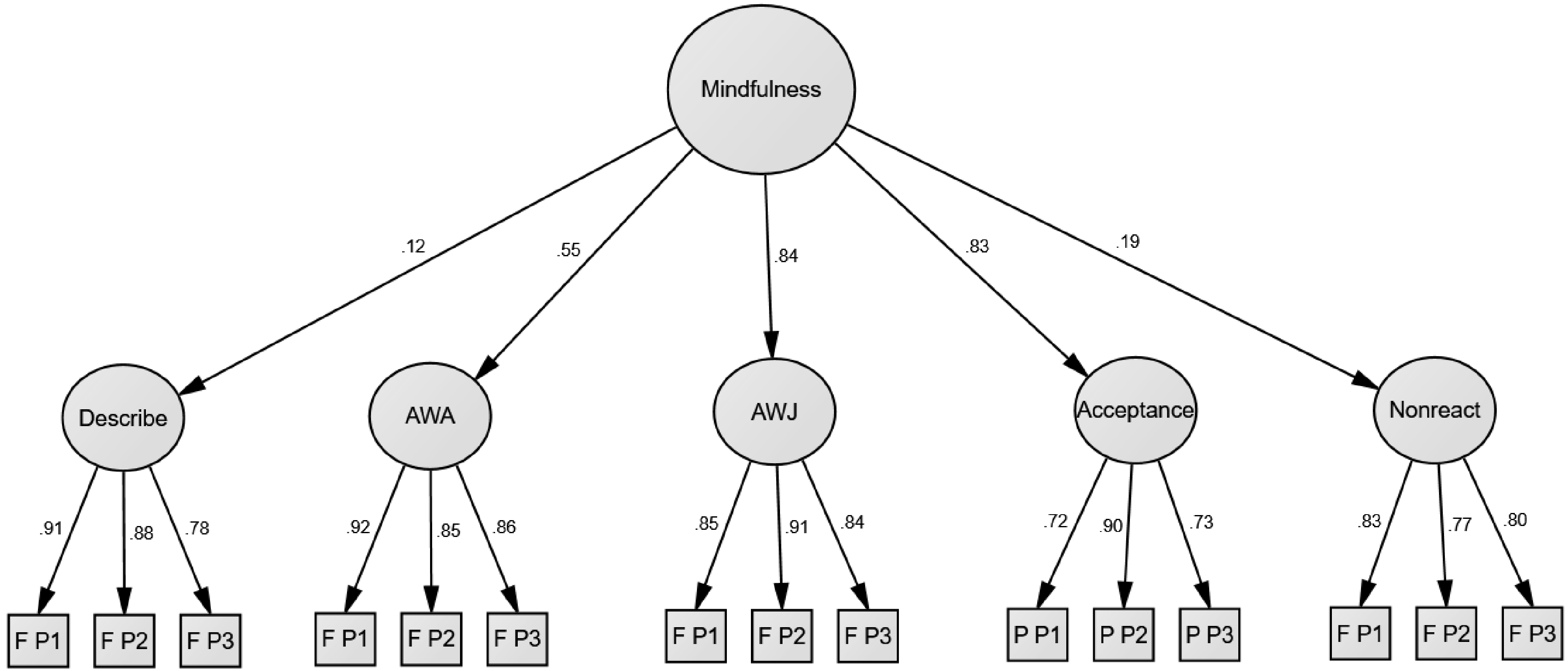


Figure 3.3. Study 2: Results for Joint Confirmatory Factor Analysis of the Five Facet Mindfulness Questionnaire (Baer et al., 2006; minus the Observe facet), Philadelphia Mindfulness Scale (Cardaciotto et al., 2008), and Toronto Mindfulness Scale (Davis et al., 2009) in Sample 1 (*N* = 395). First-order latent variables derive from item parcels (three per facet). Error terms are omitted for visual clarity. AWA = Act with Awareness; AWJ = Accept w/o Judgment; F = Five Facet Mindfulness Questionnaire; P = Philadelphia Mindfulness Scale; P1 to P3 = Parcels 1 to 3. All standardised coefficients are significant at the .05 level.

3.3.3. Discussion

This study sought to investigate similarities, validity, and dimensionality of the FFMQ facets and PHLMS and TMS subscales. Correlations among facets or subscales within measures were examined to verify the homogeneity of the FFMQ facets and heterogeneity of the PHLMS and TMS subscales, whereas between-scale correlations were examined to verify the convergent validity of similar facets. Associations with a comprehensive mindfulness component, derived from multiple unidimensional facet or item scales, were computed to assess which of the PHLMS and TMS subscales are valid indicators of the mainstream operationalisation of mindfulness. Furthermore, the factor structure of the FFMQ was examined individually and jointly with the PHLMS and TMS subscales, in order to further examine the validity of the respective facets (or subscales) as indicators of the mainstream conceptualisation. Two distinct samples of English-speaking adults, one recruited from a major British university and the other online, were used to address these questions.

Associations among the five FFMQ facets included an atheoretical pattern of associations between the Observe facet and two other facets (Act with Awareness and Accept without Judgment), were negative in Sample 1 and non-significant in Sample 2. These results fit the general pattern of non-significant or even negative associations between FFMQ Observe and some of the other FFMQ facets seen in the literature (Baer et al., 2004, 2006; Bohlmeijer et al., 2011; Curtiss & Klemanski, 2014; Höfling et al., 2011; Tran et al., 2013). They are also in line with the non-significant or negative loadings of this facet on the latent mindfulness factor observed here and in previous research (Aguado et al., 2015; Baer et al., 2004, 2006; Cebolla et al., 2012; Christopher et al., 2012; Curtiss & Klemanski, 2014; Williams et al., 2014).

As discussed, the PHLMS and TMS depart from the bulk of measures in their bidimensional focus, each assessing two relatively narrow and mainly distinct constructs. Whereas the two TMS scores were moderately associated in both samples, the two PHLMS scales were non-significantly associated in one sample, while correlating negatively in the other. For the most part, these results are also in line with previous findings. The subscales of the state version of the TMS correlated weakly to moderately (r = .26 to .42; Lau et al., 2006) and, accordingly, were interpreted as assessing distinct, but related, latent constructs. In contrast, the PHLMS subscales were explicitly created to be orthogonal, resulting in a non-significant correlation (r = -.06; Cardaciotto et al., 2008). Only the significant negative correlation observed here for the PHLMS subscales in one sample deviates somewhat from prior observations. It may be explained by the fact that Awareness is conceptualised as a deliberate behavioural process that directs one's attention towards a restricted range of experience and, simultaneously, prevents one from being open to, and accepting of, the full range of the psychological experience (Cardaciotto et al., 2008).

The next step taken in the current study was to assess the facets' relationships between the three measures. PHLMS Awareness correlated with FFMQ Observe and Describe, whereas PHLMS Acceptance correlated with FFMQ Act with Awareness, Accept without Judgment, and Nonreact in both samples. These distinct patterns of associations with mindfulness facets are consistent with the orthogonal nature of the two PHLMS subscales and previously reported correlations with the KIMS facets (Cardaciotto et al., 2008), all of which are also measured with the FFMQ. As can be expected based on conceptual similarity and previous findings, PHLMS Acceptance had the highest associations with FFMQ Accept without Judgment, whereas PHLMS Awareness correlated most strongly with FFMQ Observe.

Both TMS subscales were associated with the FFMQ Observe and Nonreact facets, and TMS Curiosity also correlated with FFMQ Describe. Also in accordance with conceptual similarity and previous findings, FFMQ Observe was the strongest correlate of TMS Curiosity, while FFMQ Nonreact was the strongest correlate of TMS Decenter (Davis et al., 2009). However, deviating from previously reported associations (Curiosity: r = .20, 95% C.I. = .11 to .29; Decenter: r = .43, 95% C.I. = .35 to .50; Davis et al., 2009), both TMS subscales were unrelated to FFMQ Act with Awareness, and TMS Decenter was also unrelated to FFMQ Describe. Furthermore, TMS Curiosity was unrelated to the FFMQ Accept without Judgment in Sample 2 and even correlated negatively with it in Sample 1. The associations reported in the current study better illustrate the bidimensional nature and distinct conceptual meanings of the TMS subscales. Also in line with conceptual resemblance, TMS Curiosity and PHLMS Awareness correlated moderately in both samples.

Study 1 showed that a single dimension underlies the shared variance of the unidimensional facet or item measures. In the current study, correlations of the PHLMS and TMS subscales with a composite derived from these measures were significant in both samples. The majority of these correlations were within, and none above, a moderate range of .30 to .70, substantiating their conceptualisation as narrower segments of mindfulness. However, correlations for TMS Curiosity and partially PHLMS Awareness were weak, suggesting that they share relatively little variance with mindfulness. Similarly, both TMS subscales previously correlated with all mindfulness scales that were used for validation purposes: the MAAS, FMI, CAMS-R, and SMQ, but correlations were generally weaker for TMS Curiosity (r = .22 to .48) than for TMS Decenter (r = .47 to .74; Davis et al., 2009). In the development study of the PHLMS

(Cardaciotto et al., 2008), the subscales correlated weakly to moderately with the MAAS, a single-factor scale included in the mindfulness component used here.

In keeping with previous findings and the problematic pattern of associations exhibited by FFMQ Observe, the four-factor hierarchical model omitting this facet fit the data best in both samples. However, facet loadings were heterogeneous, with two facets (Describe and Nonreact) showing markedly weaker loadings on the latent mindfulness factors. As observed previously (Baer et al., 2006), the two facets with the strongest loadings in both samples were Act with Awareness and Accept without Judgment. Factor analysis identified PHLMS Acceptance as an independent factor under the global mindfulness construct, alongside the four remaining FFMQ facets

3.4. General Discussion

3.4.1. Measures of dispositional mindfulness

A single dimension explains the shared variance of the scales aligned to the mainstream conceptualisation of mindfulness, and factor loadings suggest that they all tap into the same construct, albeit to different extents. Some of the scales seem to assess different parts of the construct, notably the SMQ and MAAS, which had relatively weak correlations with the other scales. For comprehensive measurement of mindfulness, the FFMQ, KIMS, and CAMS-R seem to be the best options at present, whereas the MAAS appears to be least representative, as substantiated by its relatively narrow focus on mindful attention and awareness.

In agreement with the divergent underlying conceptualisation of mindfulness, the findings speak to the distinctiveness of the LMS from the other measures. Although the results indicate some overlap of the LMS with the other scales, as has been previously found (Pirson et al., 2012), it appears that most of the variance in its global composite score is due to a different dimension. In view of these and previous relevant results, it seems prudent for future research to explicitly and systematically differentiate between the mainstream and divergent conceptualisations of mindfulness.

Despite limitations in existing research on personality and mindfulness, some inferences can be made with relatively high confidence from the reliable findings. First, both Giluk's (2009) study and ours suggest that Neuroticism, followed by Conscientiousness, are the two strongest personality correlates of mindfulness, as conceptualised in original, Eastern perspectives. Second, although the shared variance between the Five-Factor Model and mindfulness was not assessed in Giluk's study, the magnitude of associations reported in her study are similar to ours. Dispositional mindfulness, thus, seems to share considerable variance with the FFM, which the present results indicate to be around 50%. Third, linkages of the Big Five to the mindfulness construct based on Langer's (1989) divergent conceptualisation appear to be different from those of the mainstream conceptualisation advanced by Kabat-Zinn (1994); Openness is the predominant personality dimension in this construct.

3.4.2. Facets of dispositional mindfulness

The results concerning the FFMQ, and in particular its Observe facet, fit into an increasingly observed pattern of findings that speaks to the distinctiveness of this facet. Provided that the other four FFMQ facets represent mindfulness, it would not be unreasonable to drop the Observe facet entirely, at least for non-meditating samples. The problem is that such facets compromise the validity and explanatory effects of the global composite and measure when combined with the other facets (Smith et al., 2003).

The findings confirm that the subscales of both multidimensional measures scrutinised in this study assess distinct dimensions that either overlap to a small degree

(TMS) or are completely orthogonal (PHLMS). Of particular interest, however, is that the variance of at least one subscale of each of these two measures seems largely accounted for by a different construct than that underlying unidimensional facet or item mindfulness scales. The shared variance of PHLMS Awareness and especially of TMS Curiosity with mindfulness appears to be negligible; both had insufficient loadings on, and associations with, mindfulness to be considered useful facets of the global construct. These particular findings, therefore, do not support the validity of these two subscales as indicators of the mainstream conceptualisation of mindfulness. In fairness, it is important to emphasise that neither of them is conceptualised as representing a global mindfulness factor, although each has been linked to the concept. Together with the LMS, which is also dimensionally distinct, use of the TMS Curiosity and PHLMS Awareness subscales for the purpose of assessing "mindfulness" is not empirically supported.

The findings show that the PHLMS Acceptance subscale measures a potential facet of the mainstream mindfulness conceptualisation. In particular, it measures facets akin to FFMQ Accept without Judgment. However, its associations with this corresponding FFMQ facet was not large enough to suggest equivalence, or that it measures the same attribute to a similar degree. It is possible that this subscale represents the facet partially but also incorporates manifestations of the facet not already covered in the FFMQ or similar measures.

The findings pertaining to PHLMS Awareness, and to a lesser extent TMS Curiosity, may have further key implications for the representation and measurement of mindfulness. As discussed, the PHLMS assesses Awareness orthogonally to Acceptance, whereas the conceptually similar FFMQ/KIMS Observe facet is treated obliquely to the other facets, including Acceptance. Yet, the present findings show that

PHLMS Awareness (and not just Acceptance) also correlates well with its corresponding FFMQ facet (Observe). Since research is increasingly identifying the FFMQ/KIMS Observe facet to be problematic, Cardaciotto et al. (2008) may be on the right track in assessing Awareness as a distinct dimension (from Acceptance). The similar concepts reflected in PHLMS Awareness, FFMQ/KIMS Observe, and TMS Curiosity appear to be largely distinct from the mindfulness dimension underlying most of the existing measures.

3.4.3. Limitations and future directions

Some limitations of the present study must be noted. Unlike previous studies, the conclusions regarding the linkages between mindfulness and the Five-Factor Model are based on a single sample that was also relatively homogenous. A second limitation in this respect is the exclusive reliance on a single measure of the FFM. It is possible that the Big Five Inventory used in the present study may not represent the Big Five as accurately or comprehensively as other measures used in previous studies. An updated meta-analysis addressing the limitations of Giluk's study would shed light on the validity of the present results pertaining to measures aligned to the original conceptualisation of mindfulness.

Although the results were generally similar across the two samples, it is important to note that the Sample 2 data were collected online with relatively little control over who completed the survey and how. The Sample 1 data were also collected online, but these participants were recruited via the participant pool of a university, which imposes a greater degree of control and participation etiquette. Even though the data were rigorously pre-screened to identify problem responses, some invalid or poor quality responses may always go unnoticed. At the same time, the use of two (very different) samples is a notable strength of the study, with consistency in results strengthening the inferences made.

A related limitation is the use of convenience samples, with uneven distributions of demographic factors (e.g., gender) possibly impinging on the pattern of results obtained here. Although this study focused on the common pattern of associations between the two samples, it is worth noting that the correlations exhibited some differences and seemed to be generally larger in the online sample. Demographic factors, which were beyond the scope here, warrant greater attention in future research, especially in light of mixed evidence for measurement invariance (Christopher, Charoensuk, Gilbert, Neary, & Pearce, 2009; Ghorbani, Watson, & Weathington, 2009).

The present study showed that certain subscales of the PHLMS and TMS seem to qualify as facets of a global mindfulness construct. A next logical step would be to systematically examine if these subscales have any added representational value, or even advantage, relative to the FFMQ facets. PHLMS Acceptance may occupy unique construct variance not already covered by conceptually similar or equivalent FFMQ facets. For example, PHLMS Acceptance and FFMQ Accept without Judgment could be compared in their capacity to predict specific behaviours or states relevant to mindfulness. Simultaneous comparison of facets and subscales between the various measures can be very informative in regards to optimising the representation and operationalisations of mindfulness.

3.5. Conclusions

The research presented in this chapter established that the measures based on the mainstream conceptualisation of mindfulness all seem to tap into the same construct, albeit to different degrees. It was confirmed that the Observe facet assessed by two of these measures does not represent this construct at all. Likewise, only some of the subscales of two bidimensional measures were identified as indicators of the construct, as could be expected. These results render the construct of mindfulness and, in particular, the measures used in Study 2 of the present investigation suitable for the purpose of scrutinising FB. If it serves its intended purposes, FB should identify the Observe facet as an extraneous facet. It should also reveal whether the relevant subscales of the two rogue measures (PHLMS and TMS) have any added representational value in the assessment of mindfulness. This aim was pursued in **Chapter 4**.

CHAPTER 4: Application of Facet Benchmarking in the Context of Dispositional Mindfulness

4.1. Introduction

The initial application and application of FB in the context of trait EI showed promising results (**Chapter 2**). That investigation used data from six samples completing a broad, 15-facet measure of trait EI and measures of construct-relevant criteria, which varied across the six samples. The results exposed four facets that did not account for significant variance in the criterion-based construct representations in all six samples. An additional facet predicted variance in a direction opposite to that predicted by the other facets in some of the samples, showing an atheoretical effect. In all six samples, a composite of the 10 remaining facets converged significantly better with the criterion-based construct representations than the original 15-facet composite.

The varying sets of criteria used to derive the criterion-based composites implied some degree of consistency of facet effects across different construct representations. Given an ample number of trials, no additional facets accounting for significant variance in these composites emerged. An important limitation of the study was that the criteria were neither systematically nor deliberately selected for the purpose of applying FB. Although encompassing a broad and diverse set of manifestations, some elements of the construct variance may have been underrepresented in the total set of criteria examined across samples or even entirely missing. In that case, facets related to any underrepresented construct variance would not have reached significance. One aim of the present investigation was to address this limitation, by selecting the criteria specifically for representing the variance of the target construct.

In the present investigation, FB is examined in the context of dispositional mindfulness. In contrast to the initial investigation presented in **Chapter 2**, the criteria used to derive an alternative representation of the construct variance were selected deliberately (for the purpose of applying and examining FB); they were chosen systematically to ascertain a comprehensive representation of the mindfulness construct (variance). In order to apply the method rigorously, featuring built-in replication, the same criteria for deriving the criterion-based composite were assessed in multiple samples (Study 1) and a different set of criteria was used in a cross-validation of the results on a different sample (Study 2).

In both studies, FB was applied to two multi-faceted scales: the FFMQ and KIMS, which comprise five and four facets, respectively. Even though the FFMQ is an advanced version of the KIMS and includes one additional facet, the KIMS was used for validation purposes in order to increase the certainty that the results would not be influenced by any unique and potentially confounding aspects of the FFMQ (e.g., unclear items, inaccurate representation of facets, etc.). Consistent results across measures would increase confidence in the reliability and validity of findings. The other four measures on which the FFMQ is based were used for additional validation purposes in Study 1, specifically to examine the validity of the criterion-based composite. In line with the findings obtained in **Chapter 3**, the first hypothesis was:

Hypothesis 1: FB will identify FFMQ/KIMS Observe as extraneous, by showing that it does not occupy unique variance in the criterion-based composite.

An additional purpose of Study 1 was to illustrate how the method's utility extends beyond the identification of redundant or extraneous facets from within individual measures to multiple measures, involving a few very similar facets and subscales between measures. This variation of FB can provide some insight into how much, if anything, a given measure, or even an individual facet or subscale, adds to the representation of the construct. For example, it can reveal whether conceptually identical or similar facets between measures are redundant with one another, or if each encompasses unique manifestations of the same construct element (suggesting that both are too narrow). To approach this aim, FB was reapplied jointly to the FFMQ facets and the four subscales of the two bidimensional scales (PHLMS and TMS). Although three subscales (PHLMS Awareness, TMS Decenter and TMS Curiosity) were identified as problematic in **Chapter 3** (Study 2), they were nonetheless included here with the aim of further demonstrating the efficacy of FB in distinguishing problem and valid facets. FB should be able to cross-validate the results obtained in **Chapter 3**. Thus:

Hypothesis 2: PHLMS Awareness and TMS Curiosity will emerge as problematic, also lacking unique explanatory effects on the criterion-based composite.

4.2. Study 1

The criteria were selected based on their use in previous development studies of mindfulness scales. Specifically, the variables most frequently employed as validation criteria and which can be conceptualised as proximate outcomes were used. This procedure resulted in a manageable, yet decent number of criteria, offering a fairly comprehensive representation of the construct variance. The focus was on proximate psychological outcomes rather than broader outcomes that are multiply determined (e.g., clinical and mental health criteria, such as alexithymia, depression, and anxiety) or other personality constructs (e.g., EI). Table 4.1 shows the chosen criteria, along with their occurrence in the development studies of the various mindfulness scales.

Table 4.1. *Commonly Used Validation Criteria in the Development of Mindfulness Scales*

Variables	Mindfulness scales validated against variables
Experiential avoidance	FFMQ, KIMS, CAMS–R, PHLMS
Rumination and reflection	CAMS–R, LMS, MAAS, TMS, PHLMS
Thought suppression	FFMQ, CAMS–R, LMS, PHLMS
Worry	CAMS–R, LMS
Absent-mindedness	FFMQ, TMI
Dissociative activities	FFMQ, KIMS, FMI, TMS
Absorption	KIMS, MAAS
Self-consciousness	MAAS, FMI, TMI
Positive and negative affect	SMQ, LMS
Emotion regulation	FFMQ, LMS

Note. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); CAMS–R = Cognitive and Affective Mindfulness Scale – Revised (Feldman et al., 2006); PHLMS = Philadelphia Mindfulness Scale (Cardaciotto et al., 2008); LMS = Langer Mindfulness Scale (Bodner & Langer, 2001; Pirson et al., 2012); MAAS = Mindful Attention Awareness Scale (Brown & Ryan, 2003); TMS = Toronto Mindfulness Scale (trait version; Davis, Lau, & Cairns, 2009); FMI = Freiburg Mindfulness Inventory (Walach et al., 2006); SMQ = Southampton Mindfulness Questionnaire (Chadwick et al., 2008).

4.2.1. Method

4.2.1.1. Samples and procedure

FB was applied to three samples. Samples 1 and 2 were split-halves of the first sample used in **Chapter 3**, but with an equal number of male and female students assigned to each subsample; Sample 3 was equivalent to the second sample of that chapter. A summary of their characteristics is shown in Table 4.2, whereas the recruitment and data collection procedures can be found in in the preceding chapter.

Table 4.2. *Study 1: Demographic Characteristics of Study Samples*

	Age (years)			Gender (*n*)		Ethnicity (%)				
Sample (*N*)	*M*	*SD*	Range	Male	Female	Caucasian	Asian	South Asian[a]	African	Other/mixed
1 (199)	21.9	4.3	18.0–57.2	46	153	55.3	27.1	9.0	3.5	5.0
2 (198)	21.9	5.4	18.2–55.0	46	152	55.0	29.8	7.1	1.0	7.1
3 (176)	36.7	14.4	15.7–76.2	36	140	84.1	2.8	1.7	4.5	6.8

Note. Samples 1 and 2 two are split-halves of a university student sample.
[a]Includes Pakistani, Bangladeshi, Indian, and Sri Lankan backgrounds.

4.2.1.2. Measures

All measures were based on self-report, using multiple-point response scales. Table 4.3 shows the number of items per variable and internal reliabilities across samples. All mindfulness scales and criteria showed consistently satisfactory reliability ($\alpha > .70$).

Table 4.3. *Study 1: Internal Reliabilities of Study Variables*

		Cronbach's α		
Scales/variables	No. of items	Sample 1	Sample 2	Sample 3
Mindfulness scales and facets				
FFMQ	39	.87	.83	.92
Observe	8	.76	.82	.82
Describe	8	.88	.88	.93
Act with Awareness	8	.87	.90	.91
Accept without Judgment	8	.92	.91	.94
Nonreact	7	.84	.83	.88
KIMS	39	.83	.78	.89
Observe	12	.82	.85	.86
Describe	8	.88	.88	.93
Act with Awareness	10	.79	.80	.83
Accept without Judgment	9	.88	.88	.91
CAMS–R	12	.77	.74	.83
SMQ	16	.82	.79	.87
MAAS	15	.83	.88	.88
FMI	14	.83	.83	.89
PHLMS Awareness	10	.79	.77	.84
PHLMS Acceptance	10	.82	.81	.88
TMS Curiosity	6	.85	.87	.88

TMS Decenter	7	.72	.75	.76
		Criteria		
Experiential avoidance	10	.89	.90	.93
Rumination	12	.91	.91	.95
Reflection	12	.90	.90	.92
Thought suppression	15	.89	.91	.94
Worry	16	.94	.94	.95
Absent-mindedness	25	.92	.91	.93
Dissociative activities	35	.95	.96	.96
Absorption	34	.95	.94	.95
Private Self-Consciousness	10	.77	.74	.79
Public Self-Consciousness	7	.81	.80	.80
Social anxiety	6	.80	.82	.83
Positive affect	10	.72	.81	.91
Negative affect	10	.72	.86	.92
Emotion reappraisal	6	.87	.83	—
Emotion suppression	4	.78	.77	—

Note. N = 199 (Sample 1), 198 (Sample 2), and 176 (Sample 3). Of the Sample 3 participants, only 120 completed the MAAS and FMI. Emotional reappraisal and suppression were not assessed in Sample 3 and only in 111 and 90 participants in Samples 1 and 2, respectively (these variables were removed while collecting the combined data for Samples 1 and 2, since they did not load on the same component as the other criteria). FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); CAMS–R = Cognitive and Affective Mindfulness Scale – Revised (Feldman et al., 2006); SMQ = Southampton Mindfulness Questionnaire (Chadwick et al., 2008); MAAS = Mindful Attention Awareness Scale (Brown & Ryan, 2003); FMI = Freiburg Mindfulness Inventory (Walach et al., 2006).

Mindfulness

Of the mindfulness scales described in **Chapter 3**, the following were used in the present study: FFMQ (Baer et al., 2006), KIMS (Baer, Smith, & Allen, 2004), PHLMS (Cardaciotto, Herbert, Forman, Moitra, & Farrow, 2008), TMS (trait version;

Davis, Lau, & Cairns, 2009), CAMS–R (Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2006), SMQ (Chadwick et al., 2008), MAAS (Brown & Ryan, 2003), and FMI (Walach, Buchheld, Buttenmüller, Kleinknecht, & Schmidt, 2006).

Criteria

The same measures were used as in the validation studies of mindfulness scales. However, in two cases (experiential avoidance and absorption), updated versions of the measures were used in the present study.

Experiential avoidance. This criterion was measured with the Acceptance and Action Questionnaire II (Bond et al., 2011). Higher scores indicate greater psychological inflexibility ("the rigid dominance of psychological re-actions over chosen values and contingencies in guiding action"; Bond et al., 2011, p. 678) and experiential avoidance (alteration of the form, frequency, and situational sensitivity of experiences). The scale items are responded to on a 7-point Likert scale *from* 1 (*never true*) to 7 (*always true*).

Rumination and reflection. The Rumination-Reflection Questionnaire (Trapnell & Campbell, 1999) is based as an alternative, bidimensional model of private self-consciousness. The items are based on a 5-point Likert scale, ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*), and equally distributed across the two subscales (of rumination and reflection).

Thought suppression. The White Bear Suppression Inventory (Wegner & Zanakos, 1994) comprises a single scale measuring a person's attempts to suppress particular thoughts. The items are responded to on a 5-point Likert scale, ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*).

Worry. The Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990) is a measure of worry, a dominant feature of generalised anxiety disorder. The items are responded to on a 5-point Likert scale, ranging from 1 (*not at all typical of me*) to 5 (*very typical of me*).

Absent-mindedness. The Cognitive Failures Questionnaire (Broadbent, Cooper, Fitzgerald, & Parkes, 1982) measures the frequency of mistakes people make in perception, memory, and motor function; it has also been conceptualised and used as a measure of absent-mindedness. The items have a 5-point Likert scale, ranging from 0 (*Never*) to 4 (*Very often*).

Dissociative activities. The Scale of Dissociative Activities (Mayer & Farmer, 2010) measures dissociative behaviours. Responses are given on a 5-point Likert scale and range from *Never* to *Very Frequently*.

Absorption. The Modified Tellegen Absorption Scale (Jamieson, 2005) was used in this study. It measures the disposition of getting absorbed in mental imagery, using the same items as the original scale. It differs from the original scale in using a multi-point response scale, ranging from 0 (*never*) to 4 (*very often*).

Self-consciousness. The Self-Consciousness Scale (Fenigstein, Scheier, & Buss, 1975) measures three components of self-consciousness: private self-consciousness, public self-consciousness, and social anxiety. Items are responded to on a 5-point scale, ranging from 0 (*extremely uncharacteristic* [*not at all like me*]) to 4 (*extremely characteristic* [*very much like me*]).

Positive and negative affect. These two mood dimensions were measured with the Positive and Negative Affect Schedule (Watson et al., 1988), which consists of 20 adjectives of positive or negative valence. Respondents indicate the degree to which each adjective is representative of them on a 5-point scale, ranging from 1 (*very slightly or not at all*) to 5 (*extremely*).

Emotion regulation. The Emotion Regulation Questionnaire (Gross & John, 2003) measures two distinct aspects of emotion regulation: cognitive reappraisal, which reflects people's inner experience of emotions, and expressive suppression, which reflects the behaviour linked to people's feelings. Scale items are based on a 7-point response scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

4.2.1.3. Statistical analyses

Preliminary analyses examined the appropriateness of individual criteria for their inclusion in the criterion-based composite representation of the dispositional mindfulness variance. This step was aimed at minimising the effect of arbitrariness decisions involved in excluding any criteria at Step 2, based on a pre-specified cut-off. The six unidimensional mindfulness scales, including the FFMQ, were submitted to a Principal Component Analysis, and a comprehensive mindfulness component was extracted. The two bidimensional scales (PHLMS and TMS) were not included in this component. Aside from representing specific elements of mindfulness, the two subscales of each scale are either uncorrelated or weakly associated. Criteria that did not correlate with the derived component in any of the three samples were deemed unrelated to the mainstream conceptualisation of mindfulness and excluded from further analysis.

Step 2 of FB involves extracting the principal component of the chosen criteria in each sample. If necessary, criteria showing weak loadings (< .30) on the first principal component in all samples were identified through this process and excluded from the criterion-based composite. In accordance with Step 3, the criterion-based composite was regressed separately on the FFMQ and KIMS facets in each sample, using the stepwise regression method; all facets were entered into the equation at the first step of each analysis and then successively removed (criterion: $p \geq .05$) and possibly re-entered, if their betas reached significance at later steps (criterion: $p < .05$). The same procedure was used separately for a joint analysis of the FFMQ facets and the PHLMS and TMS subscales.

Step 4 (deriving a composite of facets showing predictive effects in any of the samples used) was also performed separately for the FFMQ and KIMS facets, as well as for the combination of the predictive FFMQ facets and all four PHLMS and TMS subscales. The modified FFMQ and KIMS composites were then compared against the original scale composites in their associations with the criterion-based composite. Additionally, a composite of the predictive FFMQ facets and PHLMS and TMS subscales was compared against the original and modified FFMQ composite. Steiger's *Z* tests were computed to examine if statistically significant differences existed between these associations. At Step 5, any facets removed during the regression part of FB were correlated with the modified scale composite and criterion-based composite to identify them as redundant versus extraneous.

Follow-up analyses were conducted to examine if the criterion-based composite yields an accurate representation of the construct variance. Average bivariate correlations of the criterion-based composite with the unidimensional mindfulness scales were compared to the average intercorrelations among the mindfulness scales across samples. These analyses were conducted both with and without the FFMQ, which derives from the other scales, thus introducing potential bias, by overrepresenting its elements within the multi-scale composite.

4.2.2. Results

4.2.2.1. Preliminary analyses

Principal Component Analyses of the unidimensional mindfulness scales yielded consistent results across samples, as shown in Table 4.4. All mindfulness scales loaded on a single principal component. Naturally, loadings were highest for the FFMQ, which derives from the other five mindfulness scales included in this analysis.

Table 4.4. *Study 1: Principal Component Analyses of Mindfulness Scales*

Mindfulness scales	Sample 1 (*N* = 199)		Sample 2 (*N* = 198)		Sample 3 (*N* = 120)	
	Factor loading	Communality	Factor loading	Communality	Factor loading	Communality
FFMQ	.91	.83	.93	.86	.94	.88
KIMS	.85	.72	.84	.70	.89	.79
CAMS–R	.82	.68	.84	.70	.86	.74
SMQ	.69	.48	.60	.36	.79	.63
MAAS	.63	.40	.63	.40	.68	.47
FMI	.75	.56	.76	.58	.85	.72
% of variance	61.02		60.03		70.33	

Note. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); CAMS–R = Cognitive and Affective Mindfulness Scale – Revised (Feldman et al., 2006); SMQ = Southampton Mindfulness Questionnaire (Chadwick et al., 2008); MAAS = Mindful Attention Awareness Scale (Brown & Ryan, 2003); FMI = Freiburg Mindfulness Inventory (Walach et al., 2006).

Correlations between the derived composite and the selected validation criteria were reliable across samples in terms of statistical significance (see Table 4.5). All but three criteria (absorption, private self-consciousness, and emotion suppression) correlated significantly with the composite. As Table 4.5 indicates, two criteria (emotion reappraisal and suppression) were not administered to Sample 3. Based on a preliminary analysis of the data collected from Samples 1 and 2, it was decided to remove them from further data collection, because they did not load on the same principal component.

Table 4.5. *Study 1: Correlations between Validation Criteria and Composite of Mindfulness Scales*

Criteria	Sample 1 (N = 199)	Sample 2 (N = 198)	Sample 3 (N = 120)
Experiential avoidance	-.59***	-.63***	-.76***
Rumination	-.60***	-.51***	-.60***
Reflection	.25***	.07	.26**
Thought suppression	-.57***	-.56***	-.58***
Worry	-.53***	-.54***	-.52***
Absent-mindedness	-.59***	-.50***	-.68***
Dissociative activities	-.40***	-.45***	-.57***
Absorption	.09	.02	.01
PrSC	-.02	-.07	-.05
PuSC	-.37***	-.28***	-.31***
Social anxiety	-.40***	-.29***	-.43***
Positive affect	.34***	.30***	.54***
Negative affect	-.46***	-.46***	-.58***
Emotion reappraisal	.22*	.34**	—
Emotion suppression	-.09	-.06	—

Note. Emotional reappraisal and suppression were not assessed in Sample 3 and only in 111 and 90 participants in Samples 1 and 2, respectively (these variables were removed while collecting the combined data for Samples 1 and 2, since they did not load on the same component as the other criteria). PrSC = private self-consciousness; PuSC = public self-consciousness.
*p < .05. **p < .01. ***p < .001.

4.2.2.2. Dimensional reduction of criteria

Table 4.6 shows the results of the Principal Component Analyses conducted on the criteria in each sample after excluding variables that did not meet the inclusion requirements. The criteria shown in the table loaded on the first principal component in at least one, and in most cases all, of the samples. Criteria that did not meet these requirements and that were, thus, excluded from the analyses shown in Table 4.6 are reflection and emotion reappraisal. Both had relatively weak loadings on the first principal component (reflection: .19 and .23; reappraisal: -.20 and -.33) and much stronger loadings on a second or third component (ranging from .58 to .83) in Samples 1 and 2. Additionally, reflection had a negligible loading on the first component (.07) and a strong loading on a second component (.89) in Sample 3. Consequently, the criterion-based composite was derived from the remaining variables (i.e., those shown in Table 4.6), omitting reflection and emotion reappraisal.

Table 4.6. *Study 1: First Principal Component Loadings for Validation Criteria*

	Sample 1 (*N* = 199)		Sample 2 (*N* = 198)		Sample 3 (*N* = 176)	
Criteria	Factor loading	Communality	Factor loading	Communality	Factor loading	Communality
Experiential avoidance	.82	.72	.81	.72	.88	.80
Rumination	.80	.74	.76	.69	.85	.74
Thought suppression	.76	.60	.81	.67	.83	.69
Worry	.76	.67	.72	.65	.80	.71
Absent-mindedness	.66	.50	.71	.56	.67	.66
Dissociative activities	.68	.74	.69	.76	.71	.74
Public self-consciousness	.66	.74	.61	.69	.64	.69
Social anxiety	.62	.55	.55	.66	.68	.65
Positive affect	-.26	.91	-.17	.86	-.51	.28
Negative affect	.68	.70	.74	.67	.76	.59
% of variance	47.07		46.33		54.86	

Note. Reflection and emotion reappraisal were excluded from all samples, as they had weak loadings on the first principal component and strong loadings on a second or third component (reappraisal was not assessed in Sample 3).

4.2.2.3. Regression of criterion-based composite on mindfulness facets

Regression analysis results are shown in Table 4.7. Due to the large amount of data, the results presented for each analysis are limited to the initial and final models and only beta weights for facets retained in the final model are displayed. Ignoring the FFMQ's additional facet (Nonreact), which showed significant betas, results for the FFMQ and the KIMS were congruent. The same facet in the two measures (Observe) predicted variance in an atheoretical direction, opposite to that of the other facets in Sample 1 and 3, whereas it did not occupy any significant variance in the criterion-based composite in Sample 2. Also for both scales, the Describe facet predicted unique criterion variance in Samples 1 and 3, but not in Sample 2. The other two facets of both scales showed significant betas across samples.[3]

Extension of the FFMQ analysis by inclusion of the PHLMS and TMS subscales required a few iterations in Samples 1 and 3 until no facets with atheoretical (positive) beta weights remained at the final step. If present, any such facets were removed manually and the analysis was repeated. FFMQ Describe lost its significant effect in Sample 1, but maintained it in Sample 3. Of the additional PHLMS and TMS subscales, one subscale (PHLMS Acceptance) occupied unique variance in the criterion-based composite. PHLMS Acceptance remained the only significant predictor of these two additional scales and also replaced FFMQ Accept without Judgment as the strongest predictor in Samples 1 and 3.

Across samples, the final model comprising only the predictive facets and subscales accounted for virtually the same amount of variance as the initial model

[3] To examine possible method effects associated with a specific loading cut-off, a different criterion-based composite, comprised of all criteria submitted to the initial Principal Component Analysis at Step 2, was regressed on the facets. This alternative composite included criteria otherwise removed, based on weak loadings. However, this did not change the pattern of results in any of the three samples.

comprising all facets and subscales. This pattern is logical, because the effects of predictors in regression are additive only (non-significant predictors do not minimise the overall amount of variance explained).

Table 4.7. *Study 1: Stepwise Regression Analysis Summaries for (a) FFMQ Facets, (b) KIMS Facets, and (c) FFMQ Facets and PHLMS and TMS Subscales Predicting the Criterion-Based Composite*

	Sample 1 (N = 199)			Sample 2 (N = 198)			Sample 3 (N = 176)		
Mindfulness scale and facets	β	F	R^2_{Adj}	β	F	R^2_{Adj}	β	F	R^2_{Adj}
FFMQ (all facets)		43.68***	.52		57.97***	.59		72.23***	.67
FFMQ (final model)		50.01***	.50		95.77***	.59		85.10***	.66
Observe	—			—			—		
Describe	-.12*			—			-.13**		
Act with Awareness	-.33***			-.34***			-.25***		
Accept without Judgment	-.36***			-.47***			-.41***		
Nonreact	-.24***			-.31***			-.31***		
KIMS (all facets)		50.73***	.50		43.69***	.46		62.14***	.58
KIMS (final model)		62.71***	.48		84.56***	.46		83.17***	.58
Observe	—			—			—		
Describe	-.15**			—			-.21***		
Act with Awareness	-.31***			-.30***			-.21***		
Accept without Judgment	-.50***			-.53***			-.59***		
FFMQ, PHLMS, and TMS		32.16***	.59		34.43***	.61		60.45***	.75

Final model		64.57***	.56		77.69	.61		81.31***	.73
FFMQ describe	—			—			-.12**		
FFMQ Act with Awareness	-.29***			-.31***			-.19***		
FFMQ Accept without Judgment	-.21***			-.38***			-.18**		
FFMQ Nonreactive Stance	-.24***			-.28***			-.23***		
PHLMS Acceptance	-.34***			-.19**			-.42***		

Note. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); PHLMS = Philadelphia Mindfulness Scale (Cardaciotto et al., 2008); TMS = Toronto Mindfulness Scale (Davis et al., 2009).
**p* < .05. ***p* < .01. ****p* < .001.

4.2.2.4. Correlations of original and modified scale composites with the criterion-based composite

To examine the effect of the oppositely correlated Observe facet on the validity of the FFMQ and KIMS, modified composites were computed (i.e., by omitting the Observe items in each) and compared to the original scale composites in their association with the criterion-based composite. Correlations of these original and modified composites with the criterion-based composite are shown in Table 4.8. Comparison of these correlations showed that they were significantly higher for the modified scale composites (those excluding the Observe items) in all three samples. Furthermore, a second composite consisting of the predictive FFMQ facets and PHLMS Acceptance was also compared to the modified FFMQ composite. This composite (consisting of the predictive FFMQ facets and PHLMS Acceptance) showed correlations of -.74 (Sample 1), -.75 (Sample 2), and -.84 (Sample 3), all of which were significantly larger than those of the other modified FFMQ composite not including FFMQ Observe and PHLMS Acceptance: $Z = 3.64$, $p < .01$ (Sample 1); $Z = 2.83$, $p < .01$ (Sample 2); $Z = 5.31$, $p < .01$ (Sample 3). Comparison of these associations against the original FFMQ composite was therefore unnecessary.

Table 4.8. *Study 1: Correlations of the Original and Modified FFMQ and KIMS Composites with the Criterion-Based Composite*

	FFMQ			KIMS		
Sample (*N*)	Original	Modified	Steiger's *Z*	Original	Modified	Steiger's *Z*
1 (199)	-.59	-.69	5.49	-.48	-.67	6.46
2 (198)	-.59	-.71	5.27	-.44	-.64	5.38
3 (176)	-.74	-.79	4.64	-.66	-.74	3.74

Note. All coefficients are significant at $p < .001$; all Steiger Z values are significant at $p < .01$. "Original" refers to the scale composite of all five facets; "Modified" refers to the four-facet scale composite minus the Observe facet. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004).

4.2.2.5. Correlations of the Observe facet with the modified scale composites and the criterion-based composite

Correlations of the FFMQ and KIMS Observe facet with the modified composite are shown in Table 4.9. The correlations were either weak (Samples 1 and 2) or non-significant (Sample 3). However, for both measures, correlations with the criterion-based composite were weak and positive in Samples 1 and 2 (r = .15 to .18, $p < .05$), while non-significant in Sample 3 (r = -.09 and -.10, $p > .05$). Positive associations with the criterion-based composite imply that higher Observe scorers are lower in dispositional mindfulness than those scoring at the lower end of Observe, indicating atheoretical effects of this facet.

Table 4.9. *Study 1: Correlations of Observe Facet with the Modified FFMQ and KIMS Composites*

Sample (*N*)	FFMQ Observe	KIMS Observe
1 (199)	.15*	.18*
2 (198)	.15*	.18*
3 (176)	-.09	-.10

Note. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004).
*$p < .05$.

4.2.2.6. Follow-up analyses

Average intercorrelations among the mindfulness scales for Samples 1 to 3 were .52, .51, and .64, respectively. The average bivariate correlations of the criterion-based composite with mindfulness scales were negligibly larger or similar to the average scale intercorrelations (r = -.56, -.53, and -.63). However, upon excluding the FFMQ, the criterion-based composite showed a larger average correlation with the mindfulness scales across samples (r = -.55, -.52, and -.61), compared with the average intercorrelations among the mindfulness scales (r = .47, .44, and .59).

4.2.3. Discussion

The results were reliable across the three samples for both of the primary scales: the FFMQ and KIMS. They were also the same for these two measures, which comprise similar sets of facets—their main difference is the FFMQ's additional facet (Nonreact). Regression of the criterion-based composite on the facets from the two scales identified one facet (Observe) that showed an unexpected predictive effect in all three samples. In two samples, it predicted variance in an atheoretical direction, opposite to that of the other facets. Further, this facet was shown to compromise the convergence of the two scales with the criterion-based composite. Hypothesis 1, which predicted this facet to emerge as extraneous, is therefore supported by the data.

Examining the two bidimensional scales (PHLMS and TMS) alongside the FFMQ revealed one additional subscale that occupied unique variance in the criterion-based composite; together with the FFMQ facets that already had significant betas when examined in isolation, PHLMS Acceptance emerged as a reliable predictor across samples and as the strongest predictor in Samples 1 and 3. Specifically, it replaced FFMQ Accept without Judgment as the strongest predictor, which is not surprising, since that is the most conceptually similar FFMQ facet. On the other hand, FFMQ Accept without Judgment also remained a significant predictor across samples, indicating that neither of these two variables is redundant. This result is in line with the bivariate correlation between this FFMQ facet and PHLMS Acceptance; although higher than for any other FFMQ facet, the correlation between FFMQ Accept without Judgment and PHLMS Acceptance is not as high to suggest equivalence (Cardaciotto et al., 2008; **Chapter 3**). Each seems to encompass unique content of the mindfulness variance.

Basic psychometric analyses already identified the other three subscales (PHLMS Awareness, TMD Decenter, and TMS Curiosity) as problematic. Hypothesis 2 therefore predicted that these subscales would be unable to contribute additional construct variance and emerge as problematic (specific claims in regards to redundant vs. extraneous were deliberately avoided, since these subscales showed a mix of significant and non-significant effects in **Chapter 3**). The hypothesis was supported (for all three subscales). However, no strong conclusions regarding these particular subscales are warranted, since the relevant results derive from only one analysis, based on a single criterion-based representation of the construct variance.

The consistency of results across samples and criteria speaks to the integrity and reliability of FB. Furthermore, the results indicate that the criterion-based composite derived and used here not only covers the mindfulness construct variance quite well, but also shows good evidence of convergent validity itself, given the present state of conceptualisation and evidence. Yet, while the deliberate and systematic derivation of a single set of criteria is a step forward in demonstrating the efficacy of FB, it is still possible that the selected set of criteria, including those dropped from the analysis, do not represent the common variance of every facet. Given the study's focus on reliability (generalisability), which required the assessment of the same criteria across samples, it was not practical to administer measures of all criteria previously considered relevant to mindfulness. Study 2 was conducted to increase confidence in these results, by examining their robustness against a different set of criteria.

4.3. Study 2

FB was reapplied to the FFMQ and KIMS, using a different set of mindfulness-relevant criteria. This study relied on the criteria that were used in previous validation studies of mindfulness scales and not already administered in Study 1 of the present investigation. On the one hand, using the least commonly used criteria for this purpose may be a dangerous approach, since they are less likely to represent mindfulness as accurately as the more common criteria used in Study 1. On the other hand, confirming evidence from these criteria would yield strong support for the results obtained in Study 1 and for the efficacy of FB more generally. As for Study 1, broader criteria of a clinical and subclinical nature (e.g., anxiety, depression, life satisfaction) and personality traits (e.g., EI) were omitted.

The criteria used in this study were as follows (the respective mindfulness scales for which these criteria were used for validation purposes are reported in parentheses): curiosity (LMS), need for cognition (MAAS), self-monitoring (MAAS), overgeneralisation (CAMS–R), self-compassion (FFMQ), psychological mindedness (TMS), anticipatory mental coping (4 subscales; CAMS–R), and subjective happiness (PHLMS). The measure for one criterion, hopelessness (PHLMS), was not available free of charge and the criterion itself was somewhat ambiguous in terms of its suitability for this study. In contrast to the other criteria, hopelessness has somewhat of a clinical nature and represents particular thought content, rather than cognitive or affective processes or style. The decision was therefore made to proceed without this particular criterion.

4.3.1. Method

4.3.1.1. Participants and procedure

Participants were recruited in the same way as Samples 1 and 2 of Study 1. The sample consisted of 188 undergraduate and Master's students (87.2% female), mostly from Psychology and Linguistics. Participant ages had a mean of 21.1 years (*SD* = 5.7) and ranged from 17.9 to 74.5 years. Ethnic backgrounds were mostly Caucasian (56.4%) and Asian (31.4%). The remaining participants were from South Asian (India, Pakistan, Sri Lanka, Bangladesh; 4.3%), African (2.1%), or multi-ethnic (5.9%) backgrounds.

4.3.1.2. Measures

Mindfulness

Participants completed the FFMQ (Baer et al., 2006) and KIMS (Baer et al., 2004), as described in **Chapter 3**. Cronbach's alphas for the FFMQ were .78 (Observe), .90 (Describe), .84 (Act with Awareness), .89 (Accept without Judgment), and .87 (Nonreact). For KIMS, alphas were .79 (Observe), .89 (Describe), .81 (Act with Awareness), and .88 (Accept without Judgment).

Criteria

Curiosity. The total score of the Curiosity and Exploration Inventory-II (Kashdan et al., 2009) was used as an index of curiosity. The measure's 10 items are responded to on a 4-point Likert scale, ranging from 1 (*Very Slightly or Not At All*) to 5 (*Extremely*). Cronbach's alpha was .88.

Need for cognition. The Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984) consists of 18 items that are responded to on a 5-point Likert scale (1 = *extremely uncharacteristic of me*, 5 = *extremely characteristic of me*). Cronbach's alpha was .88.

Self-monitoring. The Self-Monitoring Scale-Revised (Lennox & Wolfe, 1984) measures people's "ability to modify self-presentation" and "sensitivity to expressive behaviour of others". It has 13 items rated on a 6-point scale, ranging from 0 (*certainly, always false* to 5 (*certainly, always true*). Cronbach's alpha was .87.

Overgeneralisation. Overgeneralisation, the disposition to generalise from individual failures to one's overall self-worth, was assessed with a 4-item subscale of the Attitudes Towards Self Scale (Carver, Voie, Kuhl, & Ganellen, 1988). The items are based on a 5-point Likert scale (1 = I agree a lot, 5 = I DISagree a lot). Cronbach's alpha was .79.

Self-compassion. The Self-Compassion Scale – Short Form (Raes, Pommier, Neff, & Van Gucht, 2011) consists of 12 items suitable for assessing the global construct and showing near-perfect correlations with the full form. The items are responded to on a 5-point scale, ranging from 1 (*Almost never*) to 5 (*Almost always*). Cronbach's alpha was .88.

Psychological mindedness. The Psychological Mindedness Scale (Conte, Ratto, & Karasu, 1996) was originally designed to provide insight into clients' suitability for psychodynamic therapy, but it has since been more broadly conceived of as measuring a person's capacity for tolerating psychological distress (Shill & Lumley, 2002). Using a 4-point Likert scale (*strongly agree* to *strongly disagree*), respondents indicate the extent to which 45 self-descriptive items represent them. Cronbach's alpha was .87.

Anticipatory mental coping. The Measure of Anticipatory Mental Processes (Feldman & Hayes, 2005) assesses two productive and two unproductive strategies for coping with future stressful events. The four subscales and their internal consistencies in the present study are (Feldman & Hayes, 2005, pp. 490–491): problem analysis ("active contemplation of the antecedents and meaning of future stressful situations", 5 items, $\alpha = .86$), plan rehearsal ("envisioning the steps or strategies one could use to achieve a desired outcome", 3 items, $\alpha = .65$, $\omega = .81$), stagnant deliberation ("tendency to dwell repetitively on a stressful life problem and to experience unproductive thoughts about it", 5 items, $\alpha = .78$), and outcome fantasy ("tendency to respond to potential problems by daydreaming or fantasising about desired outcomes, while ignoring details of the problem-solving process", 2 items, $\alpha = .79$). Respondents are asked to imagine an important, difficult, and stressful problem to resolve and then to indicate how often various items reflect their typical response in this kind of situation. Subsequently, they are asked to rate how well each of the items reflects their behaviour in these situations on a 5-point scale (1 = *Never true for me*; 5 = *Always true for me*).

Subjective happiness (Lyubomirsky & Lepper, 1999). The Subjective Happiness Scale consists of four items that are responded to on a 7-point Likert scale from 1 to 7. The scale anchors vary across items. Cronbach's alpha was .82.

4.3.1.3. Statistical analyses

Since the FFMQ presents a relatively broad operationalisation of dispositional mindfulness, deriving from the other five unidimensional scales, it was used in a preliminary analysis to examine correlations between mindfulness and the criteria. As in Study 1, this preliminary step aimed at identifying any individual criteria that can be excluded from the main analyses. The statistical steps of FB (2–5) were executed as in Study 1: (1) the remaining criteria were submitted to a Principal Component Analysis to identify any criteria that did not "fit" with the others in terms of their loadings on the first component; (2) the extracted first component was regressed separately on the FFMQ and KIMS facets using stepwise regression, starting with all facets in the initial model; (3) the associations between a composite of predictive facets (in Study 1 and Study 2) with the criterion-based composite were compared to those of the original scale composites; (4) finally, associations of any non-predictive facets with the modified scale composites (comprising only predictive facets) and with the criterion-based composite were examined.

4.3.2. Results and discussion

4.3.2.1. Preliminary analyses

Significant correlations between the FFMQ global score and the criteria were observed for psychological mindedness (r = .18, p < .05), overgeneralisation (r = .50, p < .001), self-compassion (r = -.54, p < .001), curiosity, (r = .27, p < .001), need for cognition, (r = .25, p < .001), self-monitoring (r = .19, p = .008), subjective happiness (r = .46, p < .001), and two of the mental anticipatory coping subscales (stagnant deliberation [r = .37, p < .001] and outcome fantasy [r = .21, p = .003]). The only criteria not correlating significantly with the FFMQ were the other two subscales of the anticipatory mental coping measures: planned rehearsal (r = .13, p = .07) and problem analysis (r = .08, p = .28). Therefore, these two subscales were excluded from further analyses.

4.3.2.2. Dimensional reduction of criteria

Results of the Principal Component Analysis of the remaining criteria are shown in Table 4.10. The factor loadings in the table pertain to the first principal component underlying the shared variance of the criteria. All except three criteria (need for cognition, self-monitoring, and psychological mindedness), which also showed relatively weak associations with the FFMQ, loaded adequately on the first component ($\lambda \geq .30$); psychological mindedness fell below this cut-off after dropping the other two criteria from the analysis. The criterion-based composite was derived from the remaining six criteria, with factor loadings for this composite shown in parentheses in Table 4.10.

Table 4.10. *Study 2: First Principal Component Loadings for Validation Criteria*

Criteria	Factor loading	Communality	% of variance
Curiosity	.53 (.45)	.70 (.32)	31.35 (44.91)
Need for cognition	.28	.75	
Self-monitoring	.24	.47	
Overgeneralisation	.77 (.84)	.84 (.71)	
Self-compassion	-.80 (-.85)	.83 (.76)	
Psychological mindedness	.31	.82	
Stagnant deliberation	.59 (.58)	.70 (.66)	
Outcome fantasy	.39 (.43)	.76 (.75)	
Subjective happiness	.76 (.74)	.71 (.67)	

Note. N = 188. Values in parentheses derive from an analysis excluding variables that did not satisfy the specified criteria for inclusion: need for cognition, self-monitoring, and psychological mindedness.

4.3.2.3. Regression of criterion-based composite on facets

Results of the stepwise regression analysis are shown in Table 4.11. Concerning the FFMQ, two facets failed to make the final step due to non-significant betas: Observe and Describe. The former was already unable to occupy variance in the expected direction in Study 1, whereas the latter had relatively weak beta weights in Study 1, compared to the other facets. For the KIMS, only the Observe facet dropped, whereas Describe remained a significant predictor. Despite using a completely different set of criteria, the results obtained for this sample are quite similar to those obtained in Study 1 for both the FFMQ and KIMS.[4]

[4] Again to examine possible method effects linked to using a specific loading cut-off, a criterion-based composite comprised of all criteria submitted to the initial Principal Component Analysis (Step 2) was regressed on the facets. Whereas the effect

Table 4.11. *Study 2: Stepwise Regression Analysis Summaries for FFMQ and KIMS Facets Predicting the Criterion-Based Composite*

Mindfulness scale and facets	β	*F*	R^2_{Adj}
FFMQ (all facets)		33.08***	.46
FFMQ (final model)		54.02***	.46
Observe	—		
Describe	—		
Act with Awareness	.22***		
Accept without Judgment	.39***		
Nonreact	.42***		
KIMS (all facets)		18.99***	.28
KIMS (final model)		23.97***	.27
Observe	—		
Describe	.14*		
Act with Awareness	.44***		
Accept without Judgment	.17**		

Note. *N* = 188. FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004).
**p* < .05. ***p* < .01. ****p* < .001.

for Observe remained non-significant, that for FFMQ Describe became significant. FFMQ Describe already showed significant effects in Study 1 and in the current study also for the KIMS, based on the same criteria.

4.3.2.4. Correlations of the original and modified scale composites with the criterion-based composite

Although the FFMQ Describe facet did not reach significance in this sample, it did show significant effects in Study 1 and, for the KIMS, also in the current study. Therefore, it was included in the modified FFMQ composite. All four correlations were significant at $p < .001$. Zero-order correlations of the original FFMQ and KIMS composites with the criterion-based composite were .61 and .46, respectively. As expected, the correlations involving the modified composites were slightly higher at .63 (FFMQ) and .50 (KIMS), although these differences were not significant for either the FFMQ, $Z(185) = 1.16$, $p > .05$, or the KIMS, $Z(185) = 1.32$, $p > .05$. While these associations did not differ significantly, the important finding is that the associations involving the modified composites are not lower. It is also important to keep in mind that the selected set of criteria has not been as commonly considered "mindfulness-relevant" as the criteria used in Study 1. It was expected that the shared variance of the criteria used in the present study would not represent mindfulness as accurately as that of the Study 1 criteria.

4.3.2.5. Correlations of the Observe facet with the modified scale composites and the criterion-based composite

The FFMQ Observe facet did not correlate significantly with the modified FFMQ composite, $r = .02$, $p = .74$. Likewise, the KIMS Observe facet did not correlate with the modified KIMS composite, $r = -.06$, $p = .44$. Furthermore, FFMQ Observe ($r = .07$, $p = .36$) and KIMS Observe ($r = .02$, $p = .75$) were both unrelated to the criterion-based composite. These results are in agreement with those in Study 1, where this facet correlated non-significantly or modestly with the modified scale composites and the criterion-based composite. These associations suggest that the Observe facet is extraneous, and the atheoretical beta weights seen in two of the regression analyses of Study 1 provide further support for this inference. In conjunction with Study 1, the current study provides strong evidence that FFMQ/KIMS Observe is not a valid and useful facet of mindfulness.

4.4. General Discussion

This chapter presents the second application of FB, which was scrutinised in two studies (four samples in total) and primarily applied to two multi-faceted measures of dispositional mindfulness, the FFMQ (Baer et al., 2006) and the KIMS (Baer et al., 2004). Two additional measures were used to examine how FB performs when applied to multiple measures, and to illustrate its potential in signifying if other measures and their subscales have any added value, or if individual measures represent the construct comprehensively. This aim also served the purpose of cross-validating results of basic psychometric research, as conducted in **Chapter 3**.

A key methodological improvement over the first application of FB presented in **Chapter 2** is that the criteria were selected deliberately and systematically, increasing the likelihood that the construct variance was represented comprehensively. Relevant criteria used in previous validation studies of mindfulness scales representing proximate affective, behavioural, or cognitive criteria were used. Hence, the chance of any valid facets not being represented in the criteria was minimal. Another major advantage over the first investigation is that the same set of criteria was administered to multiple samples in Study 1 and these results were then cross-validated on a different, also systematically selected set of criteria in Study 2. These methodological advances provided the foundation for ascertaining the method's reliability and efficacy.

Moreover, FB was applied separately to two multi-faceted measures of the same construct, based on almost identical models.

4.4.1. Summary and interpretation of results

In both studies, the same facet (Observe) was unable to occupy unique construct variance, as represented by the composite derived from the respective sets of criteria. This pattern was observed in all three samples of Study 1 and in the single sample used in Study 2. Furthermore, removal of the Observe facet from the scale composites led to an improvement in construct validity in Study 1, as shown by the larger associations of the modified scale composites with the criterion-based construct representation. In Study 2, correlations with the criterion-based composite were not significantly different between the original and modified scale composites for either the FFMQ or the KIMS, but the modified composites again showed slightly larger correlations. Considering all four samples (both studies), the Observe facet was found to share negligible variance with the modified scale composites and the criterion-based composites.

These results are in line with the findings presented in **Chapter 2**. This study identified five problem facets, which, like the Observe facet, compromised the construct validity of the measure used across samples and construct representations. Yet, the current results also depart from those obtained in the trait EI study (in a positive way). The Observe facet was found to share virtually zero variance with the modified composites or the criterion-based composites, indicating that it is mainly an extraneous facet. In contrast, the five trait EI problem facets shared variance with the modified composite in most or all of the samples, suggesting that they are redundant. The current results, therefore, provide good support that FB can spot not only redundant facets, as shown in its first application, but also extraneous facets. Although the extant psychometric approaches have some (limited) potential in identifying extraneous facets, they cannot detect them reliably and with certainty, as the current results illustrate.

The results presented here involving the FFMQ and KIMS Observe facet are also in line with validation studies of these measures. The Observe facet was shown to load weakly on the scale composite and to correlate non-significantly or even negatively with some of the other facets, with a four-factor model typically showing a better fit (Aguado et al., 2015; Baer et al., 2004, 2006; Bohlmeijer et al., 2011; Cebolla et al., 2012; Christopher et al., 2012; Curtiss & Klemanski, 2014; Höfling et al., 2011; Tran et al., 2013; Williams et al., 2014). Moreover, this facet has predicted mostly non-significant variance vis-à-vis the other facets, also showing a mix of positive and negative effects (Cash & Whittingham, 2010; Christopher & Gilbert, 2009; Vujanovic et al., 2010). Although it is apparent across the accumulation of studies conducted that the Observe facet produces very problematic results, the current investigation shows that FB can provide specific, and much more efficiently gathered, evidence for the problematic nature of a facet. Despite the accumulating evidence base, the FFMQ continues to be used in its original form (i.e., including this facet), indicating how difficult it currently is to convince researchers of the necessity for modifications to a given measure.

4.4.2. Implications

There is now quite good evidence to argue that FB has value in advancing the validity of construct representations and assessment instruments based on them. In particular, FB seems to have utility in refining psychometric measures by identifying redundant and extraneous facets, which the conventional approaches do not accomplish.

Another implication of the findings concerns specifically the representation and operationalisation of mindfulness. It seems tempting to re-examine the Observe facet included in the two multi-faceted mindfulness scales scrutinised in this study. In conjunction with previous findings of weak factor loadings and atheoretical relations to other facets, the present results suggest that this facet should be revised, replaced, or dropped entirely. To be considered a valid facet, theoretical development and further empirical research are needed to isolate and reconceptualise any mindfulness variance this facet may occupy.

The present results give fairly good indication that the Observe facet is mainly extraneous (i.e., unrelated to the construct), but somehow survived factor-analytic work, possibly because of specific (construct-unrelated) variance that is shared with some of the other facets. The findings also suggest that the mindfulness component represented in FFMQ Accept without Judgment and PHLMS Acceptance is not fully captured by either of these variables, which could be integrated or expanded. In contrast, the TMS does not seem to represent any unique variance of the dimension shared with the mainstream mindfulness scales.

4.4.3. Limitations and future directions

A limitation of the current study as well as of the initial investigation presented in **Chapter 2** is that all measures used a self-report response format. Common-method variance may have biased the results, by influencing associations between facets and criteria differentially (Matthias Ziegler et al., 2013). Although such a scenario is arguably less likely—the results are consistent with factor-analytic results and there is no prima facie evidence that the Observe facet differs from the other facets in terms of method variance—future research integrating different measurement formats for the construct and the criteria is needed to rule out potential method effects. As discussed in **Chapter 2**, this endeavour is not straightforward, since the measurement method has implications for the construct being assessed. For example, typical-performance measures tap into traits and maximum-performance measures into abilities. One plausible solution is to measure either the construct or the criteria using informant ratings from close relatives, friends, colleagues, etc. Other possibilities may involve behavioural observations, electronic diaries, and biodata.

4.5. Conclusions

All told, these very promising findings substantiate the efficacy and integrity of FB, given their consistency with previous mindfulness research and the advanced design of the current study, relative to the initial application presented in **Chapter 2**. They speak to the reliability of FB when using the same criteria across samples and even where a different set of criteria is used. Beyond demonstrating the method's efficacy in identifying redundant and extraneous facets, the findings of this investigation suggest that FB can also identify individual measures (or facets) as incomprehensive or even as completely redundant, relative to other measures (or conceptually similar facets in other measures). This information can then be used as a basis for refining multi-faceted assessment instruments, specifically by adding missing scale content in the form of facets or items. Altogether, the evidence gives good reason to pursue further applications and development of FB, and to assert its value for research and applied purposes. If proven to aid the representation and operationalisation of psychological constructs, FB can have far-reaching implications.

CHAPTER 5: The General Factor of Motivation

5.1. Introduction

Of the three broad pillars of differential psychology (personality, ability, as well as motivation and interests; Chamorro-Premuzic, 2007), motivation has been the least researched and possibly also the least assessed in non-academic contexts. For example, searching the relevant databases for tests and measures yields the fewest hits for motivation and by far the most for personality. Moreover, unifying theories and structural models are fairly well-established in the areas of personality and intelligence. The prevailing model of personality is the multidimensional Five-Factor Model, whereas those of intelligence are the Wechsler Model (Bowden, Saklofske, & Weiss, 2011; Saklofske, Hildebrand, & Gorsuch, 2000) and Cattell-Horn-Carroll Theory (McGrew, 2005), which both assume a unidimensional hierarchical structure, featuring *g* at the apex. In contrast, great diversity characterises the substantive motivation measurement literature at the conceptual and taxonomic levels (Mayer, Faber, Xu, Faber, & Xu, 2007); no single model seems to accommodate the various types of constructs linked to this area.

A fundamental omission in the motivation assessment literature that contributes to the disjointedness characterising the field is a lack of differentiation between motivation and motives. The present chapter draws a conceptual distinction between the two and proposes the existence of a General Factor of Motivation (GFM). As described in more detail below, motives reflect what a person wants to attain or the ulterior reasons for one's behaviour (e.g., money, friends, love, power, or revenge), whereas motivation concerns a person's propensity to actually (try to) pursue personal motives (interests, goals, preferences, needs, attitudes, desires, etc.) or fulfil pre-scribed roles, whatever these may be. The chapter then proceeds to a basic construct validation programme, by systematically deriving and examining a comprehensive representation and structural model of the GFM, uncontaminated by motives and based on two measures. The validation was not designed to be comprehensive, but nevertheless sufficient to warrant the application of FB to the GFM (see **Chapter 6**); questions of substantive and structural validity were examined systematically, whereas aspects of external validity (e.g., convergent and discriminant) were examined on a preliminary basis. Prior to this, a brief review of the literature concerning individual differences in, and measurement of, motivation is necessary to anchor the present work.

5.1.1. Motivation and individual differences: Conceptualisations and measurement

As more in-depth overviews and syntheses can be found elsewhere (e.g., Bernard, Mills, Swenson, & Walsh, 2005; Mayer et al., 2007), a general analysis of the key themes and conceptualisations in the motivation assessment literature is provided here. Specifically, the focus is on the core conceptual and psychometric characteristics of motivation measures as they relate to the subsequent description of the new construct.

When considering the motivation assessment literature, the term "motivation" can appear as semantically and conceptually misleading. Historically and contemporarily, personality in the academic literature has been broadly defined as *how* a person is like and does what he or she does. Motivation has been mainly used to explain *why* a person does what he or she does (e.g., Cattell & Kline, 1977; Chamorro-Premuzic, 2007), although other important aspects, notably the *to what extent* a person does something, have been acknowledged (e.g., Chamorro-Premuzic, 2007). Accordingly, the vast majority of measures claimed to assess motivation, or some aspect of it, really tap into what are motives (e.g., success, power, and affiliation, curiosity, vocational needs, and failure avoidance) or cognate constructs, such as interests, goals, and attitudes (Cattell & Kline, 1977; Mayer et al., 2007).

Examples of measures assessing a set of multiple, universally relevant motives, often containing the "big three" (Affiliation, Power, and Achievement), are Jackson's (1984) Personality Research Form, the Reiss Profile (Reiss, 2004; Reiss & Havercamp, 1998), the projective Thematic Apperception Test (Murray, 1943), Hogan's Motives, Values, and Preferences Inventory (Hogan & Hogan, 1996), and the Unified Motives Scales (Schönbrodt & Gerstenberg, 2012). Specific area measures assess multiple motives relevant to a particular context. Some of these are broader in scope, assessing motives related to work (e.g., Kanfer & Ackerman, 2000; Weiss, Dawis, England, & Lofquist, 1967), academic (e.g., Dolan, 1983), or athletic contexts (e.g., Gill, Dzewaltowski, & Deeter, 1988). Narrower area measures focus either on a very specific motivational area, such as the motivation to be physically active (e.g., Dishman, 1980) and to self-injure (Osuch, Noll, & Putnam, 1999), or on a single motive, such as the motive to approach/avoid success (Nygård & Gjesme, 1973; Zuckerman & Allison, 1976) and to be an environmentalist (Pelletier, Tuson, Green-Demers, Noels, & Beaton, 1998).

Other measures assess what Mayer et al. (2007) have termed "motivational dynamics". Examples of motivational dynamics include the distinctions of intrinsic and extrinsic motivation (e.g., Vallerand et al., 1992), mastery and performance orientation (e.g., Seifriz, Duda, & Chi, 1992), as well as approach and avoidance motivation (e.g., Elliot & Church, 1997). Similar to motives, these measures do not tap into motivation per se, but focus on the types of reasons for, and nature of, people's motivation. Further, some of these measures are intertwined with particular motivational areas. For example, Vallerand et al.'s (1992) Academic Motivation Scale assesses three different types of intrinsic motivation (knowledge, accomplishment, and stimulation) and three types of extrinsic motivation (identified, interjected, and external regulation).

Hogan and Hogan (1996) emphasised that families of concepts like motives, values, preferences, attitudes, needs, and interests are closely related and often used interchangeably; that they differ primarily in their breadth and level of abstraction. Recent empirical support for the integration of these cognate families of concepts comes from the integration of different measures into the Unified Motives Scales, a psychometric exemplar (Schönbrodt & Gerstenberg, 2012). While the mainstream motivation literature appears to be focused on these interrelated classes of constructs, measures directly and explicitly assessing motivation, without being contaminated by motives of some form, or are hard to find and have only emerged in recent years. The only of this kind are the recently developed Motivation and Energy Inventory (MEI; Fehnel, Bann, Hogue, Kwong, & Mahajan, 2004) and the Motivation and Engagement Scale (Liem & Martin, 2012). However, the former has a clinical focus on current motivational states ("the past four weeks") and the latter, although conceptually reflective of motivation, integrates motive-like concepts, such as failure avoidance, mastery orientation, and uncertainty control, as well as facets that are more purely motivational (e.g., persistence, task management, and planning).

Additional measures may fit a pure conceptualisation of motivation, as delineated in more detail in the next section. Yet, they would be concealed as measures of other constructs, possibly because their lack of focus on a particular type or domain of motives does not fit into the mainstream literature, or perhaps to emphasise the novelty of the construct being assessed. For example, Duckworth's Grit Scale (Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Quinn, 2009) consists of the factors passion and perseverance, two "non-motives" that are highly reflective of motivation. However, the word *motivation* does not appear in the development articles of this scale. Measures of narrower constructs, such as ambition or determination, may also tap into motivation, or core aspects of it.

On the flipside, some constructs have been explicitly labelled as a form of motivation (or related types of attributes) but primarily reflect a particular motive or need, rather than a facet or factor of general motivation. A prominent example is achievement motivation, which has been defined from multiple perspectives, such as goal theory, expectancy value theory, and the perhaps the most influential concept of need for achievement (a trait-like characteristic influencing behaviour across situations, related to the competition with one's own performance norms). Achievement motivation was found to be multidimensional (Matthias Ziegler, Schmukle, Egloff, & Bühner, 2010) and may be best situated within the context of needs (as the need for achievement), along with affiliation (desire for social relationships) and power (desire of influencing and controlling others). This triad of needs constitutes McClelland's acquired needs theory, which describes motivation as the acquisition of these three needs (McClelland, 1965). Overall, achievement motivation does not seem to represent a pure, let alone general, motivation construct.

The two traditional and most widespread approaches to assessing motivation and motives are psychometrically similar to those used to assess personality: questionnaires and projective measures (Mayer et al., 2007). Thus, measures of motivation and motives generally assess typical and stable patterns of behaviour, thinking, or feeling (hence the place of these concepts in differential psychology (as one of the three pillars of individual differences); they are effectively measuring traits and, thus, are conceptually similar to personality in their focus on generalised person attributes. A few exceptions, however, focus on recent or current motivational states. Such measures ask the person about behaviour in the specified period up to the presence (Fehnel et al., 2004) or use information, such as biodata, to infer what motives are driving the person at present (Emmons, 1986; Little, 1983). Measures such as these, although undoubtedly related to people's motives and general level of motivation, primarily concern motivational states.

5.1.2. Motivation ≠ motives

There are good reasons to draw an explicit conceptual distinction between motivation and motives. As discussed, motives refer to what a person wants or does, whereas motivation describes the extent of investment. For example, imagine two people who, due to similar upbringings and social backgrounds, have the same life goals but differ in their motivation to realise them. Having a certain motive does not necessarily entail motivation to pursue it. Furthermore, when we talk about our level of motivation to engage in various behaviours, we often do not specify their purpose or the context most relevant to us, nor do we imply that our ulterior motive for our behaviour is low or high. This conceptual difference has only been implicitly recognised in the analogous distinction between goal content, which resembles motives, and goal pursuit, which is more in line with motivation (Massey, Gebhardt, & Garnefski, 2008). Cybernetic Big Five Theory, which distinguishes between personality and characteristic adaptations, offers further support for the distinction (DeYoung, 2014). Certainly, one can be more motivated to engage in one type of task (e.g., schoolwork) over another (e.g., sports). It is not claimed that motives and motivation are unrelated; it is simply argued that they are conceptually distinct psychological constructs that should be measured independently.

A second reason concerns the diversity and multidimensional structure of motives and cognate constructs, such as interests, goals, and desires; research shows that multiple distinct dimensions underlie individual differences in these types of constructs (Bernard, 2009; Bernard, Mills, Swenson, & Walsh, 2008; Jackson, 1984; Reiss, 2004; Reiss & Havercamp, 1998). This finding is logical, for there is no reason for people who have a particular motive (e.g., power) to also have a distinctly different motive (e.g., affiliation). The same can be said about cognate constructs, such as interests or goals. The finding also is consistent with the plural term used for *motives* (*interests, goals, desires, etc.*) in the name of the respective scales and wider literature to refer to these categories of attributes (e.g., Chamorro-Premuzic, 2007; Shah & Gardner, 2008).

We do not speak of "motivation*s*", or when we do, we really mean motives. For instance, when we talk about our level of motivation we often do not specify its purpose or the context most relevant to us, nor do we imply that our ulterior motive for our behaviour is low or high. Also visualise those days on which we have "zero motivation" to do anything. Whatever our personal motives or domain of our behaviour (motivational area), it makes sense that a single mechanism is responsible for the extent of our investment in various aspects of life. A unidimensional conceptualisation of motivation is also supported by the unidimensionality of cognate constructs, such as "Grit" (Duckworth et al., 2007) or amotivation (Vallerand et al., 1992). Of course, this is not to say that one cannot be more motivated to engage in one type of task (e.g., schoolwork) over another (housework). Here, it is not claimed that motives and motivation are unrelated; it is simply argued that they are conceptually distinct psychological constructs that should be measured independently. An explicitly distinct conceptualisation, and especially a systematically derived representation, of people's general level of motivation is largely missing in the extant literature.

5.1.3. The General Factor of Motivation: Theory and nomological network

Given the conceptualisation and line of reasoning presented so far, coupled with evidence from cognate constructs, the existence of a GFM that is fundamentally distinct from motives is proposed. The main similarity to motives is that both overlap with personality conceptually. In that respect, the word "general" has a twofold meaning; it signifies the cross-situational nature as well as the temporal stability (i.e., trait-like nature, as distinct from state motivation) of the construct. Although it is valuable to assess motivation as it relates to specific contexts and at specific points in time, the GFM is cross-situational and, accordingly, postulated to explain much of the variance in narrower motivational areas, such as motivation for education, work, and leisure. The construct is purely descriptive in that it captures the general level of motivation, or the propensity to actually (try to) fulfil one's chosen and pre-scribed roles, and to pursue one's goals, whatever these may be.

The GFM can be conceptualised as part of the human personality (Corr, DeYoung, & McNaughton, 2013), concerning typical patterns in affect, behaviour, and cognition implicated in a person's level of investment in whatever he or she wants to do or is expected to do. It permeates multiple personality dimensions, especially those involving considerable effort, such as Conscientiousness and Neuroticism (Emotional Stability). However, another part of the construct is theorised to be specific, or unrelated to the established personality dimensions. Thus, a systematically derived set of GFM facets should yield a distinct factor in a joint factor analysis with the Big Five facets (cross-loadings are also assumed, given overlap). The GFM's theoretical basis and its integration of attributes from multiple personality dimensions can be expected to offer (unique) explanatory advantages in relevant contexts and for several important outcomes.

The role of motivation in personality, particularly in the context of the Five-Factor Model, has previously been identified and discussed by Denissen and Penke (2008). These authors adopted a multidimensional view of motivation. For example, Openness (to experience) was defined as individual differences in the activation of reward system during active cognitive processing, and Agreeableness as differences in the motivation to cooperate during resource conflicts. In our view, the same motivational mechanism runs across personality dimensions and is responsible for distinct behaviours (cooperating, being social, engaging in cognitive activity), but its activation varies between individuals across domains. In line with this view is the conceptualisation of motivation as a process of seeking (tendency to approach intellectually challenging situations) and conquering (tendency to master those challenging situations) in the context of intellect, intrinsic motivation, and curiosity (Mussel, 2013). Both these two highly interrelated constructs (r = .86) correlated with multiple dimensions of the FFM, foremost Openness/Intellect (r = .72 and .71) and Conscientiousness (r = .20 and .47).

Although its systematic examination is beyond the scope of this investigation, it is imperative to consider the GFM's nomological network further, with regard to other individual-differences constructs and outcomes. Conceptually most similar to, and perhaps highly converging with, the GFM are the aforementioned constructs of grit, engagement, amotivation, and achievement motivation, as well as other related constructs, such as goal pursuit (as opposed to goal content, which is more similar to motives; Massey, Gebhardt, & Garnefski, 2008). Perhaps at least moderately associated, but not quite as highly overlapping, are the discussed personality dimensions included in established trait taxonomies. Particularly the domains that appear to involve motivational facets, such as Conscientiousness and Neuroticism, can be expected to show considerable overlap with the GFM. At another level, the GFM should relate to several psychological outcomes. For example, depression often involves a severe lack of motivation, whereas mania is at the high extreme, characterised by excessive motivation (5th ed.; Diagnostic and statistical manual of mental disorders; American Psychiatric Association, 2013). First, however, measures based on a systematically derived, comprehensive representation of the proposed GFM are needed.

5.1.4. The present research

The purpose of the present investigation is to systematically develop and validate a comprehensive operationalisation of the GFM, one that is uncontaminated by motives. To this end, two forms of the "Drive", intended for research purposes in the general adult population, were developed and validated in a series of five studies. The scale construction approach was deductive in that, in Study 1, a comprehensive corpus of human traits, the International Personality Item Pool (IPIP; Goldberg et al., 2006), was examined by two independent raters for theoretically relevant facets of the GFM. Moreover, the identified facets were evaluated for their face validity by a panel of six raters in Study 2, in order to increase confidence in them. The scale construction approach was also inductive, because no *a priori* factor structure could be readily determined on theoretical grounds. In the first instance, the derived representation was examined empirically for its reliability, homogeneity, and factor structure in Study 3, using the corresponding IPIP scales. Study 4 was aimed at cross-validating the derived model by means of a short form, which was validated simultaneously. On the basis of available data, Study 5 presents preliminary evidence for the external validity (i.e., convergent, discriminant, and criterion validities) of the measure. Studies 3 to 5 were each based on the data of three or four different samples.

5.2. Study 1: Facet Selection

The IPIP was used as a platform for identifying relevant facets. Broadly speaking, the same methods are used to measure personality and motivation constructs, and the GFM, in particular, permeates the space of personality conceptually. A discussed earlier, personality concerns *how* a person is like and behaves, whereas motivation describes *what* a person does (does not) and to *what extent* he/she does something in terms of commitment, sacrifice, effort, etc. To varying degrees, the specific facet-level traits assessed with major personality inventories, and hence the IPIP, fit with a cross-situational and temporally stable conceptualisation of motivation, representing typical behaviour, thoughts, feelings, etc.; they are not exclusively representative of personality. For example, the facet of adventurousness is linked to the Big Five domain of Openness, but, it also seems to reflect motivation in that, although a person may be very open to experience, he or she may lack the energy to embark on an adventure. As a second example, consider the facet of diligence included under the Conscientiousness domain in the HEXACO model of personality. Besides being Conscientious, the diligent individual needs motivation to actually approach tasks in a diligent manner. The key point is that facet-level traits comprise of substantial specific variance unaccounted for by higher-order personality dimension. Some of that variance (from multiple personality domains personality) relates to motivation conceptually and is theorised to constitute the GFM.

Additional reasons are practical and concern the comprehensive representation of the GFM, adequate measurement of its constituent facets, as well as uses and target populations. The IPIP provides an exhaustive set of individual-differences traits (245 at the time of this study), presumably representative of the entire GFM domain, to choose from; it integrates the facets of the major personality inventories that have appeared in the scientific literature and, therefore, extends beyond the facets used to represent and operationalise the Big Five personality traits. For all of these facets, the IPIP provides a uniform and validated set of items that are in the public domain and can be readily used. Another advantage concerns the equivalent target population (the general adult population) and uses (predominantly basic research) envisioned for the two Drive forms advanced in this investigation.

Importantly, the focus was exclusively on the narrowest trait level (facets) and not on the broader, higher-order dimensions (e.g., traits such as Openness or Agreeableness were not considered) in order to ascertain a similar level of abstraction, or breadth, among the facets, as well as to ascertain their basic categorical similarity (Costa Jr. & McCrae, 1995). Moreover, facets are conceptually more neutral than higher-order personality dimensions in that they are not exclusively representative of personality; they seem to incorporate variance of other sources than personality, such as motivation.

5.2.1. Method

5.2.1.1. Raters

The two raters (one male, one female) came from different academic backgrounds. One of the raters was a PhD student in Psychology, with specific academic knowledge in the area; the other was manager of a university library. The academic credentials of this professional were a Masters in Library and Information Studies and a BA (Hons) in History, including a minor in Psychology.

5.2.1.2. Materials and procedure

The nature of the construct, as explained in the introduction, was discussed among the raters. Subsequently, the raters read through the entire set of IPIP facets (245) in order to identify relevant attributes reflective or indicative of motivation in a generic sense, independent of a person's specific circumstances, states, goals, desires, interests, motives, etc. The raters completed the task independently on a computer, simply by clicking "yes" or "no" to indicate the relevance of each facet. As a reminder, the following instruction was presented on the screen: "Which of the following characteristics are reasonable indicators of motivation? Motivation here refers to a person's general level of motivation, not tied to any particular circumstances (context or time), goals, desires, interests, etc." Raters were allowed to look up the meaning of any adjectives in a general language dictionary; definitions were not provided.

5.2.2. Results and discussion

The inter-rater reliability was strong (Cohen's $\kappa = .88$) and 24 facets were identified as relevant by either or both of the two raters. Disagreement only occurred in five instances: prudence, temperance, self-efficacy, self-control, and self-confidence. To ascertain that the preliminary set of facets would be overinclusive, rather than underrepresentative and, therefore, that no relevant facets would be omitted (see Simms & Watson, 2007; Smith, 2005a), facets identified as relevant by only one of the two raters were retained for Study 2, along with those on which both agreed. This preliminary set of facets is shown in Table 5.1.

Table 5.1. *Studies 1 and 2: General Motivation Facets Selected from IPIP and Their Endorsements by Six Independent Raters*

	Number of endorsements		
Facet	"Yes – motivation"	"Maybe – not sure"	"No – not motivation"
Self-confidence	4	1	1
Temperance	3	2	1
Zest/enthusiasm/vitality	6	0	0
Valor/bravery/courage	3	2	1
Liveliness	4	2	0
Insight	4	1	1
Initiative	5	1	0
Diligence	5	1	0
Deliberateness	5	1	0
Competitive	4	2	0
Experience-seeking	5	1	0
Generates ideas	5	1	0
Prudence	2	2	2
Resourcefulness	4	1	1
Self-control	3	3	0
Ind./persev./persis.	5	1	0
Activity-level	6	0	0
Adventurousness	4	1	1
Self-discipline	4	1	1
Achievement-striving	5	1	0
Hope/optimism	4	2	0
Competence	3	1	2
Self-efficacy	4	1	1
Joyfulness	2	4	0

Note. IPIP = International Personality Item Pool; Ind./persev./persis. = Industriousness/perseverance/persistence.

5.3. Study 2: Face Validity of Facets

The derived set of facets was scrutinised by a separate panel of raters. This step was taken to further validate these facets in terms of their conceptual relevance and, secondly, because a more lenient selection strategy was employed in Study 1.

5.3.1. Method

5.3.1.1. Raters

The panel comprised five PhD students and one psychology post-doc in Psychology (three female, three male), who were of diverse cultural backgrounds (Turkish/British, two British, Canadian/Lebanese, Italian, and Canadian) and based either in the UK, Italy, or Canada. Three of them had a research and applied psychology focus, whereas the other three were primarily oriented towards academic research and academics.

5.3.1.2. Materials and procedure

As in Study 1, the panel members were asked to rate these attributes independently in terms of whether they are indicative of a person's general level of motivation, regardless of his or her specific circumstances, states, goals, desires, interests, motives, etc. Since the raters were dispersed geographically, they completed the task online, with all instructions presented in written form. They were asked to rate each of the 24 attributes using the following options: "Yes – motivation", "Maybe – not sure", or "No – not motivation". The instructions referred the raters to an electronic dictionary in case they needed to look up the meaning of any facets they were not sure about. Verbatim instructions are shown in Appendix 2.

To further illustrate the meaning of the target construct, the six raters were shown two conceptual definitions of motivation that reflect the target construct very well. These were: (a) "Desire and energy in people to be continually interested and committed to a job, role or subject, or to make an effort to attain a goal", as defined in the BusinessDictionary ("Motivation," n.d.-a), and (b) "general desire or willingness of someone to do something", as defined in the Oxford Dictionaries ("Motivation," n.d.-b). The rationale for showing these additional definitions was to facilitate understanding of the concept and minimise the risk of presenting a single definition that is too restrictive or biased in some way.

5.3.2. Results and discussion

Table 5.1 contains the endorsements of the six raters for each of the facets. There were no facets for which none of the raters said "Yes". "Yes" was the most frequent selection by the raters in almost all instances. The only exception was "joyfulness" for which four raters said "maybe – not sure" and two said "Yes". The "No – not motivation" option was typically selected by one or two raters or none at all. In only one case (self-control), three raters selected this option (as many as those who said "Yes"). Given these endorsements, none of the 24 attributes identified by either or both of the initial two raters were dropped at this stage; all were retained for the psychometric evaluation presented in Study 3.

5.4. Study 3: Internal Reliability, Homogeneity, and Factor Structure

This study examined the validity of the 24 derived facets empirically and describes the empirical steps in the development of the measure. Since modified versions of the corresponding IPIP scales were used to measure the 24 facets (see

Measure and procedure subsection), this version will be referred to as the Drive:IPIP.[5] The first step was to examine the internal reliability of the modified scales. Subsequently, the homogeneity of the set of facets was examined and, simultaneously, the relevance of individual facets was assessed. Eventually, the facets' factor structure was explored in two samples, using Exploratory Factor Analysis, whereupon the extracted models were compared in a third sample, using Confirmatory Factor Analysis. Two samples were used for exploratory purposes to minimise the risk of sample-specific effects impinging on the structural model and because an additional test of the ensuing model was executed in Study 4. Last, the internal reliability was re-examined for the "surviving" facets.

5.4.1. Method

5.4.1.1. Participants

All sample descriptions in this investigation (Studies 3–5) concern the valid cases only, whereas "bad cases", such as drop outs, were removed prior to data analysis from two of the three samples used in this study. Cases were deleted from the datasets if participants had completed the respective measures partially or skipped numerous items, presumably due to reasons other than any plausible discomfort caused by the items (isolated missing items were deemed acceptable). Furthermore, participants who clearly did not take the survey seriously (e.g., giving the same responses throughout or using offensive language in the demographics section) and those who had unrealistically fast completion times in samples that completed the measures electronically.

[5] The Drive:IPIP scales and their respective are in the public domain, available at ipip.ori.org. All rights and ownership reside with those who have developed the items and generously made them available for research. In keeping with the intended purpose of the IPIP items, the Drive:IPIP should be considered a public-domain measure of the GFM and is primarily intended for research purposes.

Specifically, participants who did not spend a minimum average of two seconds per item were removed. Based on an overall inspection of the datasets (e.g., missing items, response style, non-numerical answers, etc.), a two-second cut-off was deemed suitable for screening out the majority of remaining problem responses, without removing any valid cases.

An Online sample was recruited using a mix of procedures. First, a recruitment notice was posted on recruitment platforms for academic research in psychology (e.g., onlinepsychresearch.co.uk, callforparticipants.com) and on a general commercial recruitment platform for online academic research (findparticipants.com). Third, a recruitment notice was posted on twitter by a prominent, non-academic authority and writer in the area of motivation. The total sample size was 362 participants (78.2% female), who had a mean age of 33.49 years (*SD* = 13.5, range = 15.9–73.1). Ethnic backgrounds of the participants were 76.1% Caucasian, 4.5% African, 3.6% Chinese, 3.9% South Asian (Bangladesh, India, and Pakistan), and 11.9% other or mixed. Educational qualifications obtained were distributed across GCSE/O or similar (5.8%), A Level or similar (19.9%), BA/BSc or similar (34.4%), MA/MSc or similar (19.6%), MBA (3.6%), PhD (2.5%), and other (14.1%). At the time of the study, 33.1% were enrolled in full-time and 8.8% in part-time education; 43.6% were working full-time and 18.0% part-time. As an incentive and token of appreciation, participants were entered into a price draw for one of several gift vouchers.

It was also possible to use the data from the U.S.-based Eugene-Springfield community sample, which were provided by the Oregon Research Institute. The portion of the sample with complete data for the relevant IPIP scales comprised 496 adults (41.9% male) with a mean age of 50.7 years (*SD* = 11.8, range: 20 to 83 years). Virtually all of the participants were Caucasian (98.8%). The remaining were either

Asian American, Hispanic, or other. The most prevalent educational levels were "some college" (28.5%), "post-college education" (25.4%), "college graduate" (20.5%), "high school graduate" (9.3%), and "some post-college education" (9.1%); others were "vocational/technical schooling" (5.7%) and "not graduated from high school" (1.4%). At the time of study, participants were working full-time (44.6%) or part-time (14.9%), retired (19.6%), homemaker (9.0%), unemployed (3.1%), or did not specify their current role (8.1%).

A British sample (*N* = 241, 80.1% female) was recruited via the divisional subject pool of a major British university. Most participants were full-time students (88.0%) and currently studying for a BA/BSc or similar degree (62.7%) or MA/MSc or similar degree (22.8%); other qualifications pursued included MBA, PhD, and other (each less than 3%). Relatively small proportions of the sample studied part-time and worked full-time or part-time (each less than 10%). Although the age range was 18.0 to 74.1 years, the mean age of the sample was 22.4 years (*SD* = 6.7) and 93% were below the age of 30. Participants were predominantly of Caucasian (49.8%) and Chinese (32.4%) ethnic backgrounds; the remainder came from South Asian (Pakistan, India, Sri Lanka; 3.3%), African (0.8%), and other or mixed backgrounds (13.7%). Many of the participants were eligible for course credit and all were entered into a prize draw for gift vouchers.

5.4.1.2. Measure and procedure

The data from the Online and British samples were collected via an electronic survey. After providing demographic information, participants in these two samples completed the IPIP items corresponding to the selected facets. The complete set of items can be found on the IPIP website (http://ipip.ori.org/), by following the scales corresponding to the selected facets. Since several of the IPIP items are used to measure multiple facets, 33 items in total were removed from the facet scales used in this study. In order to maintain a conservative number of items per facet (Velicer & Fava, 1998), 11 items included in two or more of the relevant IPIP scales were retained for facets that otherwise would be comprised of only four items or less (without allowing any item to represent multiple facets). In total, 148 items were used to represent the 24 facets. This set of items can be found in Appendix 3, organised by facets. As shown in Table 5.2, the number of items per scale ranged between 5 and 10.

Administration instructions for the Online and British samples were adapted from the generic version shown on the IPIP website in order to fit the nature of the construct assessed. The standard IPIP response scale, ranging from 1 (*very inaccurate*) to 5 (*very accurate*), was used in all samples. Administered electronically, the order in which items appeared in the Online and British samples was randomised across participants. In case participants would skip any items by accident, missing responses were highlighted to these two samples upon completing the survey.

The Eugene-Springfield community sample completed the comprehensive set of IPIP items as part of the data collection conducted by the Oregon Research Institute.

Table 5.2. *Study 3: Internal Reliabilities of IPIP Facets and Corrected Facet-Total Correlations in Study Samples*

	Online sample (*N* = 362)		Eugene-Springfield community sample (*N* = 496)		British sample (*N* = 241)	
Facet (no. of items)	Cronbach's α	Corrected facet-total correlation	Cronbach's α	Corrected facet-total correlation	Cronbach's α	Corrected facet-total correlation
Self-confidence (5)	.78	.74	.70	.68	.70	.64
Temperance (9)	.77	.61	.73	.49	.67 (.78)	.54
Zest/enthusiasm/vitality (9)	.82	.78	.79	.62	.78	.76
Valor/bravery/courage (10)	.70	.56	.75	.51	.78	.50
Liveliness (8)	.85	.72	.82	.57	.83	.66
Insight (7)	.76	.64	.71	.53	.75	.64
Initiative (5)	.81	.74	.74	.54	.81	.68
Diligence (5)	.70	.80	.66 (.78)	.61	.69 (.80)	.74
Deliberateness (6)	.69 (.79)	.49	.60 (.75)	.39	.61 (.74)	.49
Competitive (5)	.66 (.79)	.76	.64 (.78)	.74	.59 (.76)	.76
Experience-seeking (5)	.69 (.81)	.48	.61 (.77)	.37	.65 (.79)	.37
Generates ideas (5)	.85	.45	.78	.54	.81	.54
Prudence (7)	.75	.54	.69 (.79)	.52	.71	.52
Resourcefulness (6)	.75	.85	.75	.73	.69 (.81)	.77
Self-control (9)	.71	.29	.67 (.78)	.12	.71	.19

Ind./persev./persis. (7)	.81	.76	.80	.64	.77	.67
Activity-level (6)	.79	.81	.75	.56	.79	.74
Adventurousness (7)	.73	.42	.66 (.77)	.34	.78	.35
Self-discipline (5)	.77	.77	.70	.67	.70	.71
Achievement-striving (6)	.71	.74	.64 (.77)	.64	.75	.70
Hope/optimism (8)	.81	.75	.71	.62	.70	.74
Competence (6)	.77	.88	.65 (.78)	.74	.64 (.78)	.80
Self-efficacy (7)	.78	.84	.72	.73	.70	.82
Joyfulness (7)	.86	.74	.76	.58	.79	.68

Note. Where Cronbach's alpha is low (< .70), McDonald's omega is given in parentheses. IPIP = International Personality Item Pool; Ind./persev./persis. = Industriousness/perseverance/persistence.

5.4.1.3. Statistical analyses

Corrected item-total correlations were examined at the facet level in order to screen out any obviously unrelated or “tangentially relevant” facets, based on low facet-total correlations ($r < .30$) in all three samples. Subsequently, facets were submitted to a Principal Component Analysis in the Online and British samples, which were used for exploratory purposes, due to their greater diversity than the Eugene-Springfield community sample. Preceding the Exploratory Factor Analysis by a Principal Component Analysis, as applied here, is coherent with psychometric theory (Clark & Watson, 1995; Cortina, 1993) and had a twofold purpose: (a) to verify that the facets share a common dimension (i.e., the hypothesised GFM) that accounts for a decent portion of variance (i.e., unidimensionality), and (b) to further assess the relevance of individual facets, possibly dropping any that are primarily related to distinct dimensions. This second purpose served to minimise the possibility of maintaining a potentially overinclusive representation, given the rather liberal selection procedures employed in Studies 1 and 2. In order for facets to be dropped, they had to show higher loadings on components other than the first (representing the GFM) in both samples.

Following Principal Component Analysis, a combination of Exploratory Factor Analysis and Confirmatory Factor Analysis was used to generate a structural model. As a first step, Exploratory Factor Analysis, using principal axis factoring with oblique rotation (Promax method, delta = 4), was executed to examine the facets for shared first-order factors. Factor loadings of less than .30 were suppressed. This analysis was performed on the Online and British samples, and the ensuing models were compared in the Eugene-Springfield community sample, using Confirmatory Factor Analysis. As specified and used in **Chapter 3**, the fit criteria were: GFI ≥ .90, CFI ≥ .93, NFI ≥ .90, and SRMR ≤ .09.

On theoretical grounds, the right to make modifications at the Confirmatory Factor Analysis stage, particularly concerning error covariances, was reserved here. Improving model fit by adjusting correlated errors is defensible if there are solid theoretical reasons for doing so (Hooper, Coughlan, & Mullen, 2008), which is the case here. Although the facets were derived to represent the GFM specifically, they share variance linked to higher-order personality dimensions and, therefore, can be expected to correlate due to dimensions other than the GFM. Hence, the addition of error covariances on the basis of fit indices is defensible.

The last step was to examine the internal reliability at the global composite level, based on the best-supported model. Facet instead of item scores were used to compute the internal reliabilities, because the uneven number of items across facets may lead to inaccurate coefficients.

5.4.2. Results and discussion

5.4.2.1. Internal reliability and facet-total correlations

Internal reliabilities and corrected facet-total correlations for all three samples can be found in Table 5.2. Cronbach's alphas were generally acceptable for each facet in at least one of the four samples. However, where alphas were low (> .70), McDonald's omega was computed (as described in Stone et al., 2013) as an additional estimate of internal reliability (shown in parentheses in Table 5.2), since it tends to be higher and more accurate than alpha, which is a lower bound to, and often an underestimate of, internal reliability (Revelle & Zinbarg, 2009; Stone et al., 2013). Inspection of corrected facet-total correlations suggests that all except one of the facets are linked to the GFM; self-control showed low correlations (< .30), with particularly low correlations (≤ .15) in two of the three samples. This facet was excluded from further analyses.

5.4.2.2. Principal Component Analysis

Results are shown in

Table 5.3 for both the Online and British samples. In both samples, 19 of the 23 facets had their strongest loadings on the first of four components that emerged in both samples, based on Eigenvalues greater than 1. Although all facets loaded at least to some degree on the first component, four facets (prudence, deliberateness, experience-seeking, and adventurousness) had stronger loadings on the second ensuing component than on the first component in both samples, indicating that they primarily represent a distinct dimension than the one targeted. Given the overinclusive selection procedures used in Studies 1 and 2 and the multidimensional background of the derived set of facets, it was decided to drop these four facets from the preliminary representation and not include them in further analyses. Also, the second component showed similar loadings between the two samples, suggesting that it represents a reliable dimension. Components three and four had unsystematic loadings between the two samples, presumably representing “noise”. Two facets in the Online sample (generates ideas and insight) and one facet in the British sample (temperance) had a higher loading on one or more additional components. However, to maintain a conservative approach to deleting facets and minimise the chance of dropping valid facets, these three facets were provisionally retained.

The variance explained by the first component was similar across the two samples, 60.0% and 53.4%, respectively. The fact that this component did not explain most of the variance in the facets is not unexpected, because its constituent facets derive from multiple, largely distinct personality domains. In spite of being multidimensional from a personality perspective, this set of facets seems to share variance distributed across personality domains, a single dimension theorised to represent the GFM. Importantly, these results yield preliminary support for the proposed GFM, with all of the theoretically derived facets loading onto the same component.

Table 5.3. *Study 3: Principal Component Loadings of IPIP Facets in the Online and British Samples*

	Online sample (*N* = 362)				British sample (*N* = 241)			
Facet	1	2	3	4	1	2	3	4
Self-confidence	.77				.68		.33	
Temperance	.61	-.44	.36		.54	-.49	.47	
Zest/enthusiasm/vitality	.81				.79			-.32
Valor/bravery/courage	.61	.30			.56	.35		.33
Liveliness	.75			-.37	.70			-.46
Insight	.68	.51			.69	.49		
Initiative	.78				.73	-.35		
Diligence	.84		-.34		.78	-.41	-.32	
Deliberateness	.49	-.64	.32	.39	.50	-.54	.46	
Competitive	.81		-.31		.81			
Experience-seeking	.52	.62			.41	.72		
Generates ideas	.50	.50		.37	.59	.44		.40
Prudence	.55	-.64		.34	.53	-.62	.39	
Resourcefulness	.87				.79			
Ind./persev./persis.	.79	-.40			.71	-.46		
Activity-level	.85				.79		-.33	
Adventurousness	.45	.60		.31	.39	.68		
Self-discipline	.80	-.31			.75	-.42		
Achievement-striving	.79		-.42		.75		-.41	
Hope/optimism	.77		.38		.76			
Competence	.90				.84			
Self-efficacy	.86				.84			
Joyfulness	.77		.44	-.31	.72		.34	-.37
% of variance	53.69	12.92	6.36	4.92	47.99	15.17	6.98	5.56

Note. Factor loadings of < .30 are omitted from the table. IPIP = International Personality Item Pool; Ind./persev./persis. = Industriousness/perseverance/persistence.

5.4.2.3. Exploratory Factor Analysis

Analyses were conducted on the 19 remaining facets. Kaiser's criterion of Eigenvalues greater than 1 and an unambiguous scree plot both indicated three factors, whereas a parallel analysis extracted five statistically significant factors (scree plot and parallel analysis output are provided in Appendix 4). However, with the number of factors fixed to five, one factor had only two facet loadings of .30 or higher. In view of this result, a four-factor solution was instead tested. At this point, all four factors were characterised by at least three loadings, but three facets (self-efficacy, temperance, and resourcefulness) showed critical cross-loadings ($\lambda > .30$) that were within .20 of each other; these three facets did not load distinctly on a single factor. Although somewhat controversial, dropping items with cross-loadings (within .20 specifically) is not uncommonly used as a rule-of-thumb for dropping variables in Exploratory Factor Analysis (Nunnally & Bernstein, 1994). . Critical and undifferentiated cross-loadings indicate that a facet may be relatively broad in scope and redundant with several other facets across factors. Facets with cross-loadings complicate model fitting in Confirmatory Factor Analysis. Removing these facets did not change the number of factors based on parallel analysis, but one factor identified by merely two facets appeared, and a scree plot and Eigenvalues continued to support a three factor solution. Consequently, a three-factor solution was accepted for the Online sample, the results of which are shown in Table 5.4. The three factors explained 79.4% of the variance in the facet scores and showed intercorrelations of moderate strength.

Table 5.4. *Study 3: Pattern Matrix for Promax Three-Factor Solution Extracted from IPIP Facets and Factor Correlation Matrix in the Online Sample*

	Factor loading		
Facet	1	2	3
Diligence	1.02		
Ind./persev./persis.	.96		
Initiative	.92		
Self-discipline	.90		
Achievement-striving	.85		
Activity-level	.70		
Competence	.69		
Competitive	.62		.33
Joyfulness		1.12	
Hope/optimism		.82	
Liveliness		.82	
Zest/enthusiasm/vitality		.72	
Self-confidence		.62	
Insight			.95
Generates ideas			.88
Valor/bravery/courage			.46
Eigenvalue	9.65	1.83	1.22
% of variance	60.32	11.44	7.62
	Factor correlations		
Factor 1	—		
Factor 2	.66	—	
Factor 3	.56	.62	—

Note. *N* = 362. Factor loadings of < .30 are omitted from the table. IPIP = International Personality Item Pool; Ind./persev./persis. = Industriousness/perseverance/persistence.

As was observed in the Online sample, analyses conducted on the 19 remaining facets extracted three (based on the scree plot and Kaiser's Eigenvalue criterion) and five factors (based on parallel analysis) from the British sample data. Scree plot and parallel analysis results are included in Appendix 4. The promax rotated pattern matrix showed that the fifth factor was again unfeasible, since it was identified by only two facets. Consequently, a four-factor solution was tested. Although a minimum of three facets per factor emerged, three facets showed critical and undifferentiated cross-loadings: temperance (two cross-loadings), competence, and self-efficacy (both one cross-loading). With these three facets taken out of the analysis, the fourth factor became exclusively comprised of secondary cross-loadings. When testing a three-factor solution instead, two additional facets with critical and undifferentiated cross-loadings emerged: competitive and resourcefulness. As shown in Table 5.5, their removal resulted in a clean three-factor solution, which continued to be supported by the scree plot and Eigenvalues. The three factors mirror those obtained in the Online sample, with the exception that competence and competitive were removed in the process. They also explained a similar amount of variance in the facet scores (77.3%) as they did in the Online sample and, again, showed intercorrelations that were moderate in size and not critically high.

Aside from supporting a three-factor solution, both samples identified the facets of self-efficacy, temperance, and resourcefulness as problematic, due to their undifferentiated cross-loadings. These particular facets are, perhaps, also least conceptually reflective of the GFM, compared to the 10 other facets. For these reasons, the decision was made to drop the three facets at this stage and exclude them from further analysis.

Table 5.5. *Study 3: Pattern Matrix for Promax Three-Factor Solution Extracted from IPIP Facets and Factor Correlation Matrix in the British Sample*

	Factor loading		
Facet	1	2	3
Diligence	.98		
Self-discipline	.91		
Ind./persev./persis.	.86		
Initiative	.84		
Achievement-striving	.81		
Activity-level	.69		
Joyfulness		1.04	
Liveliness		.86	
Zest/enthusiasm/vitality		.66	
Hope/optimism		.65	
Self-confidence		.58	
Insight			.88
Generates ideas			.87
Valor/bravery/courage			.55
Eigenvalue	7.48	2.18	1.16
% of variance	53.45	15.55	8.30
	Factor correlations		
Factor 1	—		
Factor 2	.56	—	
Factor 3	.42	.60	—

Note. *N* = 241. Factor loadings of < .30 are omitted from the table. IPIP = International Personality Item Pool; Ind./persev./persis. = Industriousness/perseverance/persistence.

5.4.2.4. Confirmatory Factor Analysis

The two very similar models extracted from the Online and British samples were tested in the Eugene-Springfield community sample. One facet, activity-level, showed elevated residual covariances with the other facets, with several in a range of 2 to 4. Consequently, it was decided to remove this facet from both models, also in view of its relatively abstract conceptual nature. The model derived from the Online sample (minus activity-level) yielded poor initial fit, $\chi^2(88) = 787.41$, $p < .001$, CFI = .83, GFI = .80, NFI = .82, SRMR = .09. Although the model obtained from the British sample (minus activity-level) also did not fit the data adequately, fit indices were generally higher, $\chi^2(64) = 538.16$, $p < .001$, CFI = .86, GFI = .85, NFI = .84, SRMR = .09. Moreover, by means of a cautious examination of modification indices and addition of a limited number of error covariances, fit for this second model could be improved to a reasonable level, $\chi^2(59) = 311.34$, $p < .001$, CFI = .92, GFI = .91, NFI = .91, SRMR = .08 (a similar attempt for the first model was unsuccessful). No paths were low enough to consider additional facets for deletion; the lowest pointed from the second factor to initiative at .47. Consequently, this model was accepted and used as a basis for Study 4. The results are shown in Figure 5.1.

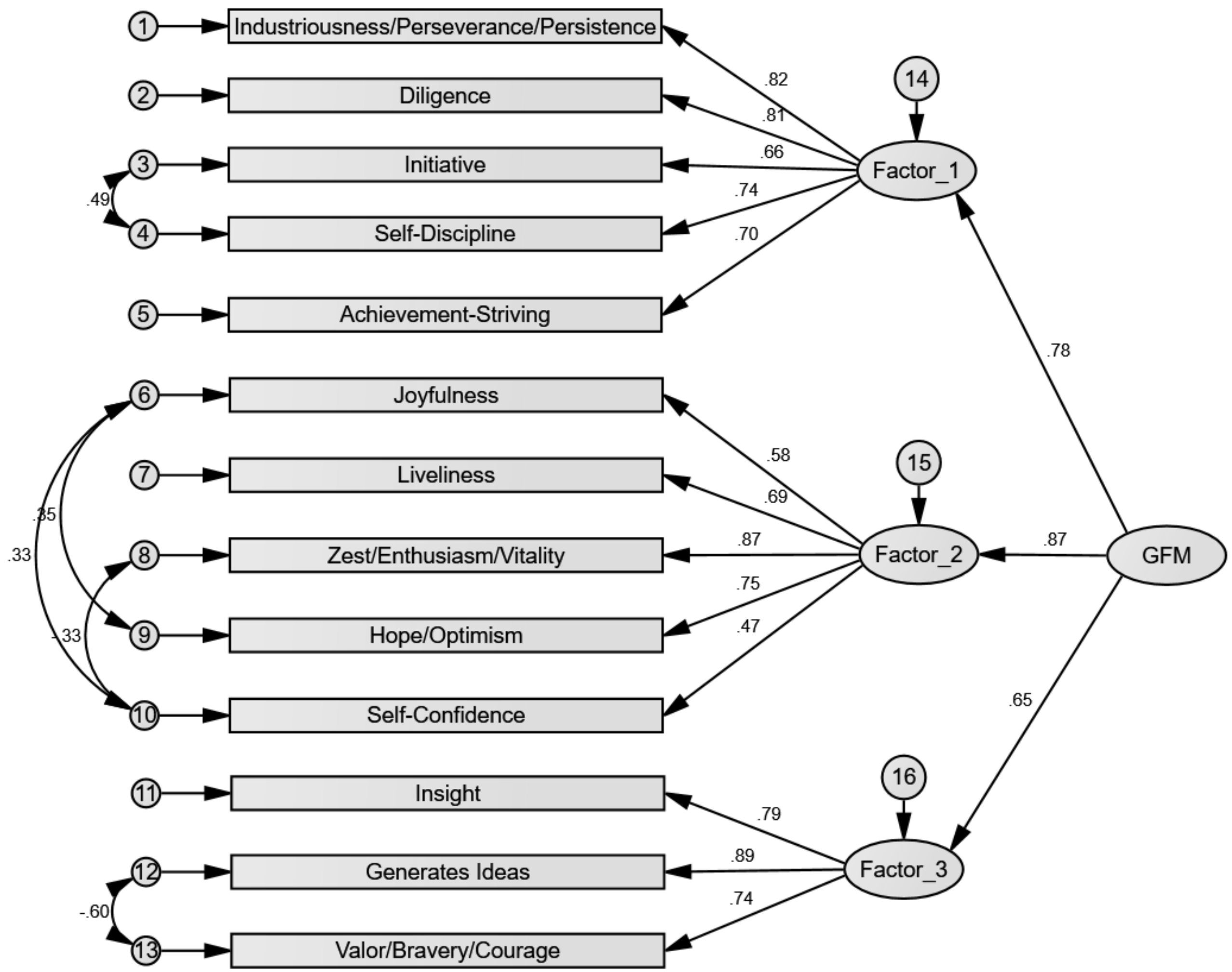


Figure 5.1. Study 3: Confirmatory Factor Analysis results for the three-factor model, uncovered from the Drive: International Personality Item Pool version in the Eugene-Springfield community sample (*N* = 496). GFM = General Factor of Motivation.

As in the Online and British samples, factor intercorrelations were moderate at .56 (Factors 1 and 2), .35 (Factors 1 and 3), and .53 (Factors 2 and 3; $p < .001$ for all) in the Eugene-Springfield community sample. These similar patterns of associations (across samples) indicate that the differences in facets between the models extracted from the Online and British samples do not affect the integrity of the factor scores substantially.

Both Exploratory Factor Analysis and Confirmatory Factor Analysis results support a hierarchical three-factor structure. The first factor comprises the facets of industry/perseverance/persistence, self-discipline, diligence, initiative, and achievement-striving; it represents the behavioural and executive manifestations of motivation and could be labelled "Vigour" or perhaps "Behavioural Motivation". The second factor comprises the facets joyfulness, hope/optimism, liveliness, zest/enthusiasm/vitality, and self-confidence; it represents the affective manifestations of motivation and could be labelled "Passion" or "Affective Motivation". The third factor consists of generates ideas, insight, and valor/bravery/courage; it represents the cognitive manifestations of motivation that could be labelled "Fortitude" or "Cognitive Motivation". Overall, these factors and their respective facets represent the "ABCs" of individual differences (Revelle, Wilt, & Condon, 2011), indicating that motivation has affective, behavioural, and cognitive manifestations that all need to be considered in an attempt to assess the construct comprehensively.

5.4.2.5. Composite Internal Reliability

The internal reliability of the global composite, based on facet scores, was consistent across the three samples: Cronbach's alpha was .95 (Online sample), .91 (Eugene-Springfield community sample), and .94 (British sample).

5.5. Study 4: Factorial Validity and Short Form Validation

Study 4 further scrutinised the validity of the structural model advanced in Study 3, by means of a short form of the measure (the Drive:S). Specifically, it was examined if a different way of measuring the facets would yield comparable reliability, homogeneity, and replicate the three-factor structure obtained with the Drive:IPIP. The Drive:S was simultaneously validated. Given study aims and practical reasons, direct estimates (i.e., single-item rating scales) were used to represent each of the facets on the Drive:S. Single-item scales are an efficient and valid way to measure specific attributes (Burisch, 1984) and were themselves used as items of a broader, multi-faceted construct. Research comparing multi-items and single items measures has not revealed empirically observable differences in construct validity and methods variance (Gardner, Cummings, Dunham, & Pierce, 1998). Relatively recent research suggests that single-item scales are particularly valid in instances where positively worded Likert items are used (Alexandrov, 2010) and when the underlying construct is homogenous (Loo, 2002), both of which is the case here. Moreover, the effects of two different response scales for these facet estimates were examined (see Measures subsection).

5.5.1. Method

5.5.1.1. Participants and procedure

The Online and British samples of Study 3 completed the Drive:S items after the IPIP items. However, due to missing items on the Drive:S, which could not be compensated by other items, the effective samples sizes were 302 and 181, respectively. An additional and distinct sample (N = 142, 61.3% female) was recruited from Norway via Qualtrics Sample Finder. This Norwegian sample also completed the Drive:S along with the Drive:IPIP, using the same procedures as described in Study 3. The mean age of this sample was 47.6 years (*SD* = 7.7) and participants reported an average of 33.8 years (*SD* = 9.5) of spoken English. The vast majority of the participants (96.5%) were of Caucasian descent. Highest educational qualifications were distributed across High school or similar (24.6%), BA/BSc or similar (22.5%), A Level, IB, or similar (19.7%), MA/MSc or similar (14.1%), MBA (9.9%), PhD (4.9%), and other (4.2%). Few participants (9.9%) were still enrolled in part- or full-time education; approximately half the sample (47.9%) were full-time workers and another 21.1% worked on a part-time basis. This sample was financially compensated for their participation.

5.5.1.2. Measures

The Drive:S uses direct estimates of each of the Drive facets, with facet labels functioning as items. The Online and British samples completed the items on a visual analogue scale with an electronic slider. This scale ranged from 0% to 100% and prompted participants to rate themselves in comparison to other people of similar age. The Norwegian sample completed the items on a 7-point Likert scale, ranging from 1 (*very little*) to 7 (*very much*). The purpose of varying the scale format was to examine, and to some extent control for, measurement effects linked to any particular scale format. Administration instructions for these facet ratings were kept as similar as possible to the item ratings but were modified to accommodate the different scale and response format. The measure including instructions can be found in Appendix 5.

The Drive:IPIP scores in the Norwegian sample showed good internal reliability (Cronbach's $\alpha = .93$).

5.5.1.3. Statistical analyses

The model derived in Study 3 was used as the input model in this study, since the IPIP scales (multiple-item rating scales) can be expected to represent the facets somewhat more accurately than the facet estimates (single-item rating scales) used in the current study. Internal reliability of the Drive:S composite was assessed using Cronbach's alpha, along with corrected item-total correlations to ascertain the empirical stability of the facets. Confirmatory Factor Analysis was executed as in Study 3, but only on the Online and British samples; the Norwegian sample was not used for this purpose, as it did not exceed the very minimum number of five cases required per model parameter (Bentler & Chou, 1987). Lastly, bivariate correlations between the Drive:IPIP and Drive:S were examined.

5.5.2. Results and discussion

5.5.2.1. Internal reliability and item-total correlations

The internal reliability of the Drive:S estimates was consistent across samples: Cronbach's alpha was .91 (Online sample), .89 (British sample), and .89 (Norwegian sample). Corrected item-total correlations were consistently acceptable (> .30), ranging from .44 for self-discipline in the British sample to .75 for zest/enthusiasm/vitality in the Online sample. The results attest to the internal reliability of the Drive:S and the validity of the individual facets. They also suggest that the use of Likert versus visual analogue scale format has little influence on internal reliability. Both seem to yield scores of adequate and comparable reliability.

5.5.2.2. Confirmatory Factor Analysis

When tested in the Online sample, the three-factor model, as tested in the Eugene-Springfield Community Sample using the Drive:IPIP (including error covariances), reached good overall fit (without any further modification), $\chi^2(59) = 183.58$, $p < .001$, CFI = .95, GFI = .93, NFI = .93, SRMR = .05. Three of the five error covariances were replicated; that between Joyfulness and Self-Confidence was non-significant, whereas that between Initiative and Self-Discipline was reversed in direction. Path coefficients and covariances for this model are shown in Figure 5.2. Factor loadings on the GFM were generally similar to those observed for the Drive:IPIP, and all three factor intercorrelations were moderate at .55 (Factors 1 and 2), .55 (Factors 1 and 3), and .63 (Factors 2 and 3; $p < .001$ for all), consistent with those reported in Study 3.

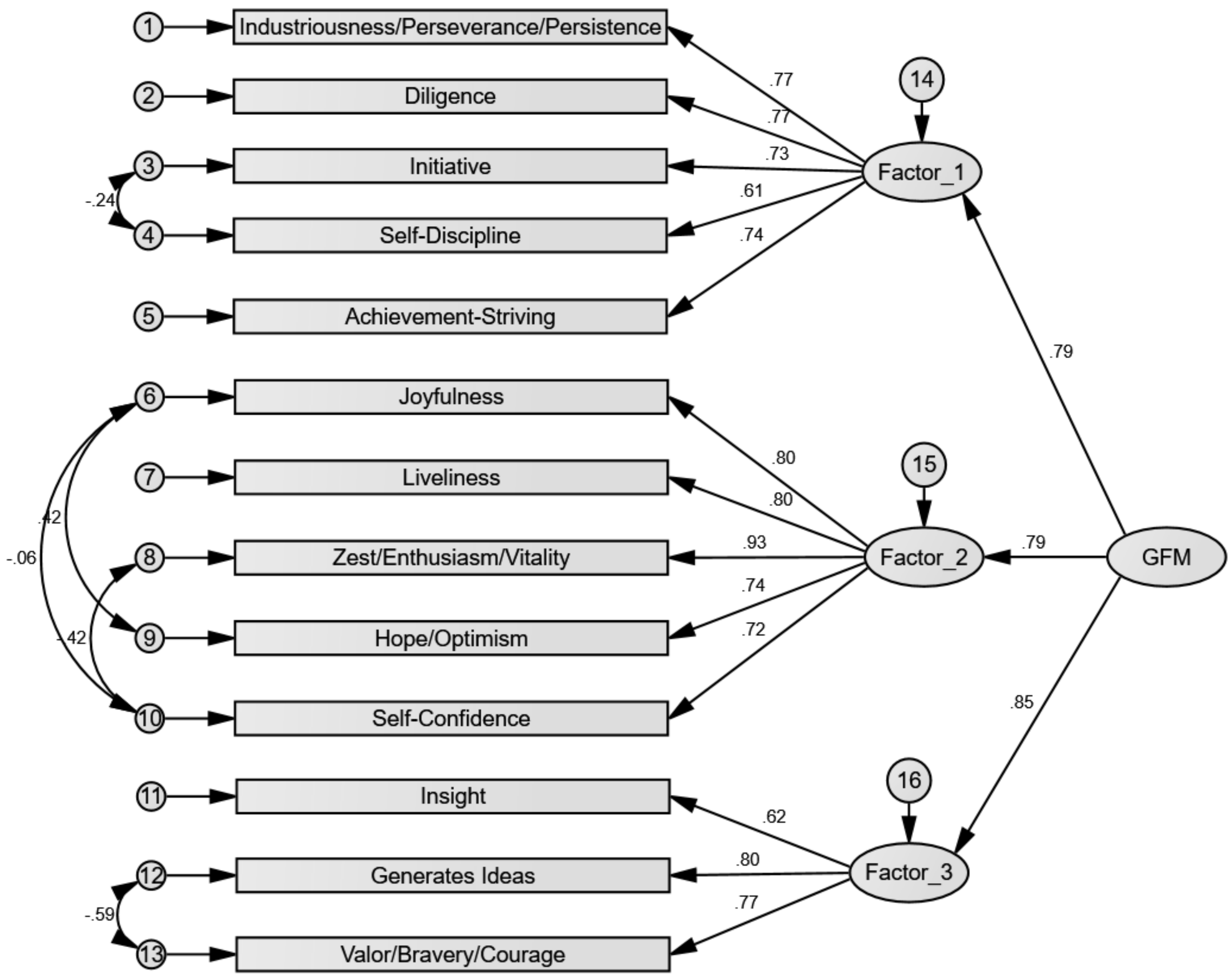


Figure 5.2. Study 4: Confirmatory Factor Analysis results for the three-factor model, recovered from the Drive: Short in the Online sample (N = 351). GFM = General Factor of Motivation.

In the British sample, model fit was adequate and comparable to that seen in the Online sample, $\chi^2(59) = 155.63$, $p < .001$, CFI = .94, GFI = .90, NFI = .90, SRMR = .08. In this instance, the five original error covariances were relatively and consistently weak. Only that between Joyfulness and Hope/Optimism was significant and of the same direction observed for the Drive:IPIP in the Eugene-Springfield Community Sample. Zest/Enthusiasm/Vitality and Self-Confidence also showed a significant covariance, but in a direction opposite to that seen initially. Path coefficients and covariances are shown in Figure 5.2. Factor intercorrelations in this sample were comparable to those in the Online sample at .45 (Factors 1 and 2), .48 (Factors 1 and 3), and .59 (Factors 2 and 3; $p < .001$ for all).

Altogether, the results obtained for the Drive:S from two different samples support the three-factor structure, comprised of five, five, and three facets, respectively. Only one of the five error covariances seen for the Drive:IPIP was consistently replicated for the Drive:S, while three of them were replicated in one of the two samples. However, the presence of all five covariances did not render the models "unfit" to the respective data. The covariances between Initiative and Self-Discipline and between Joyfulness and Self-Confidence, where significant, seem to represent sample-specific random error, possibly related to measurement effects (the response scale differed for the Drive:IPIP). The other, more reliable covariances point to systematic relationships, although of magnitude generally not comparable to the respective factor loadings of the facets concerned. They are also difficult to explain, since their relationships are negative.

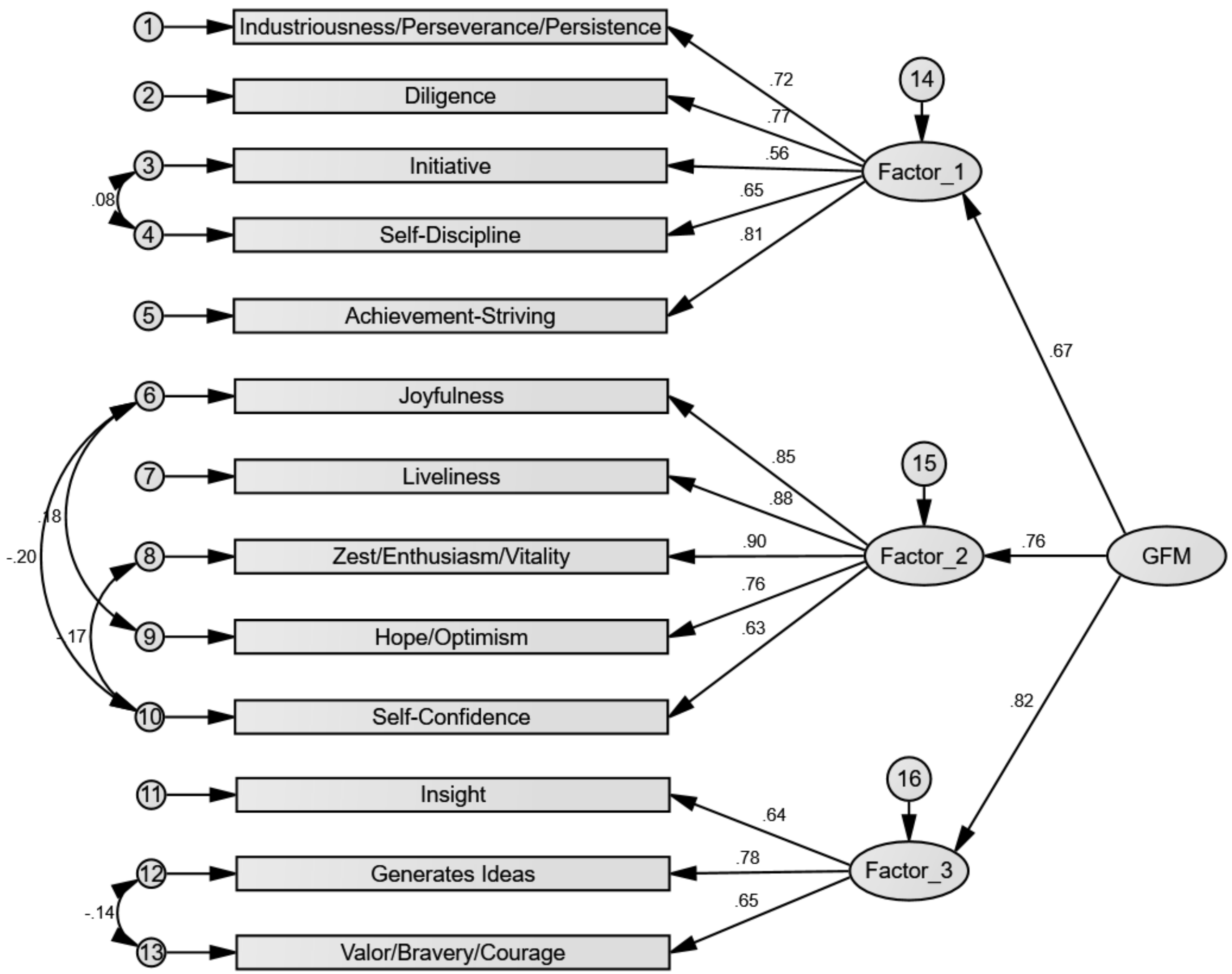


Figure 5.3. Study 4: Confirmatory Factor Analysis results for the three-factor model, recovered from the Drive: Short in the British sample (*N* = 233). GFM = General Factor of Motivation.

5.5.2.3. Drive:S and Drive:IPIP correlations

Bivariate correlations between the Drive:S and Drive:IPIP composites were similar in the Online (r = .81, p < .001) and Norwegian samples (r = .83, p < .001), but somewhat weaker in the British sample (r = .73, p < .001). Although consistently strong in magnitude, the less than "perfect" coefficients observed here are not surprising. Concerning the Drive:S, facets were measured on the basis of a proxy method of measurement (direct estimates), which may be more susceptible to measurement error than using a representative set of items. The overall results in this study are highly similar to the Study 3 findings and certainly encouraging enough for the Drive:S to qualify as a brief measure of the GFM.

5.6. Study 5: External Validity

This study presents preliminary evidence for the external validity of the Drive:IPIP and Drive:S, based largely on available data. Specifically, additional data gathered from the samples in Studies 3 and 4 were used to explore its convergent validity (hypothesised: correlations with state motivation, Conscientiousness, Ambition, Service Potential, Clerical Potential, and Managerial Potential), discriminant validity (hypothesised: non-significant or weak correlations with Agreeableness, Service Orientation, and School Success, as well as a distinct factor in Five-Factor Model personality space), concurrent validity (hypothesised: correlations with global self- and other-perceptions of motivation, depression, dietary health behaviour, exercise, and work avoidance), and incremental validity (hypothesised: significant explanatory effects on depression, dietary health behaviour, exercise, and work avoidance, controlling for personality). Some of the analyses involved informant data on the Drive:S to examine self- and informant agreement.

5.6.1. *Method*

5.6.1.1. Samples and variables

Convergent validity was examined by means of associations of both Drive forms with a state measure of motivation, completed by the British and Norwegian samples (while state motivation is situation- and time-dependent, stable individual differences in the GFM should also play a role). Since this measure was added to the data collection from the British sample with a slight delay, only 175 participant of this sample (72.6%) completed the measure. On the basis of available data, correlations between the Drive:IPIP and related personality constructs (Conscientiousness, Ambition, Service Potential, Clerical Potential, and Managerial Potential) were examined in the Eugene-Springfield community sample to provide additional insight into the measure's convergent validity. The roles of Conscientiousness and Neuroticism were emphasised in the introduction, while ambition is synonymous with motivation (and related constructs). All three constructs should show sizable correlations with the Drive:IPIP. Service, clerical, and managerial potential are compound constructs that involve some degree of motivation, next to a host of other attributes; they should show at least significant correlations with the measure.

Discriminant validity was examined in this sample by correlating the Drive:IPIP with conceptually less similar personality constructs, representing a person's interpersonal style (Agreeableness, Service Orientation, Reliability, and School Success), rather than motivation. School Success is a scale measuring the degree to which a person enjoys academic activities and values educational achievement for its own sake (Hogan & Hogan, 1992); it is more in line with motives and related constructs. The Drive:IPIP should show weaker associations, if any, with these three constructs.

The concurrent validity of both Drive forms was examined in the Online, British, and Norwegian samples, using global perceptions of motivation as the criterion. Additionally, subgroups of the Online (n = 46, 16 male) and British (n = 101, 28 male) samples nominated a person who they thought would know them well enough to provide global perceptions of motivation and complete the Drive:S as informants. Informants were friends, family members, or romantic partners, who had known them for an average of 13.6 years (SD = 11.3) and 11.5 years (SD = 8.8), respectively. Since these informant data we obtained several months after collecting the participant data, the perceptions provide a gauge for the predictive validity of the two Drive forms. On the other hand, informant data on the Drive:S were used to examine the measure's cross-informant validity, as well as to expand on its convergent validity.

Concurrent validity was further examined in the Eugene-Springfield community sample for the Drive:IPIP, with an extension to incremental validity. On the basis of availability and conceptual relevance to the GFM, the selected criteria were depression, dietary health behaviour, exercise, and work avoidance, most of which are universally important and necessitate motivation, albeit perhaps to different degrees. Incremental validity analyses were conducted on the Eugene-Springfield community sample criteria and controlled for the explanatory effects of higher-order personality dimensions, which can be expected to explain unique variance these criteria. As discussed, the personality overlaps conceptually and operationally with the GFM, as operationalised here. Therefore, controlling for their explanatory effects is important in order to demonstrate the value of the proposed construct. For work avoidance, additional analyses controlled for general and occupational personality traits assessed with a second personality measure, given their relevance to the criterion.

5.6.1.2. Measures

British and Norwegian samples: State motivation

Participants completed the MEI (Fehnel et al., 2004), which measures three highly interrelated factors (Mental Energy, Social Motivation, and Physical Energy), which were also combined to give a total motivation score. The measure consists of 27 items with some variations in the response scale. Most of the items are measured on a 6- or 7-point Likert scale (frequency type), ranging from either *Never* to *Every day or nearly every day* (e.g., “During the past 4 weeks, how often did you feel enthusiastic when you began your day?”), *Never* or *None of the time* to *All of the time* (e.g., “During the past 4 weeks, how often did you avoid social conversations with others?”), or *Never* to *At least 7 times a week* (e.g., During the past 4 weeks, how often did you engage in recreational activities or hobbies?). Another six items are “to what extent” questions and have a 5-point Likert scale, ranging from *Not at all interested* to *Extremely interested* (e.g., “During the past 4 weeks, to what extent were you interested in learning or trying new things?”). Internal reliabilities (alphas) are included in Table 5.6 for both samples.

Table 5.6. *Study 5: Descriptive Statistics and Bivariate Correlations of the Drive:IPIP and Drive:S with MEI Scales in the British and Norwegian Samples*

Variable	*M*	*SD*	Skewness	Kurtosis	Drive:IPIP	Drive:S	Mental Energy	Physical Energy	Social Motivation	Total score
				British sample (*N* = 175)						
Mental Energy	35.77	9.99	-0.55	1.00	.57	.46	(.87)			
Physical Energy	20.47	7.06	-0.27	-0.23	.42	.38	.56	(.82)		
Social Motivation	27.20	7.82	-0.27	-0.16	.36	.31	.25	.44	(.83)	
Total score	83.44	19.39	-0.60	1.01	.59	.50	.82	.83	.69	(.90)
				Norwegian sample (*N* = 142)						
Mental Energy	43.00	11.45	-1.34	2.00	.61	.45	(.91)			
Physical Energy	22.63	8.30	-0.52	-0.23	.54	.42	.71	(.86)		
Social Motivation	24.96	8.07	-0.62	0.29	.61	.52	.63	.66	(.81)	
Total score	90.78	24.10	-1.01	1.19	.68	.54	.91	.88	.85	(.94)

Note. Cronbach's alphas for the MEI scores are presented in parentheses along the diagonal. Correlations are mostly significant at $p < .001$. Drive:S response scale differed between the two samples (see Method section). Drive:IPIP = Drive: International Personality Item Pool version; Drive:S = Drive: Short; MEI = Motivation and Energy Inventory (Fehnel et al., 2004).

Online, British, and Norwegian samples: Motivation perception

An additional item ("motivation") was administered with the facet estimates, but in order to be used separately as a criterion in this study. Hence, the Online and British samples gave their global motivation perception on a percentage-based visual analogue scale, while the Norwegian sample used a 7-point Likert scale, as described and explained in Study 4.

Eugene-Springfield community sample

Personality. The NEO Personality Inventory–Revised (NEO-PI-R; Costa Jr. & McCrae, 1992) and Hogan Personality Inventory (HPI; Hogan & Hogan, 1995) were used as measures of personality traits. Variables used from these two measures were the Big Five domains and 30 facets of the NEO-PI-R as well as the seven primary scales (Adjustment, Ambition, Sociability, Likability, Prudence, Intellectance, and School Success) and six occupational scales (Service Orientation, Stress Tolerance, Reliability, Clerical Potential, Sales Potential, Managerial Potential) of the HPI. The five NEO-PI-R domains and facets comprise a total of 240 items (48 per domain, 8 per facet), which are responded to on a 5-point Likert scale, ranging from *Strongly Disagree* to *Strongly Agree*. In contrast, the HPI uses true-or-false items, ranging in number from 14 to 37 for the primary scales and from 14 to 67 for the occupational scales. It was not possible to compute internal consistency reliabilities for these two measures, since items scores were not available in the obtained datasets. However, both measures generally produce scores of high internal reliability (Costa Jr. & McCrae, 1992; Hogan & Hogan, 1992), and adequate levels of internal reliability have been reported in previous publications using these data, as listed on the IPIP website.

Depression. Depression was measured using a modified 24-item version of the Center for Epidemiologic Studies Depression Scale (Radloff, 1977), which was extended by a few items. This measure was developed specifically to assess depressive symptomatology in non-clinical populations. Respondents indicate how frequently they experienced a range of depressive symptoms during the past week (e.g., "I had a poor appetite"). Items were responded to on a 5-point Likert scale, ranging from 1 (*not at all past week*) to 5 (*most or all of the time*). Internal reliability was adequate (Cronbach's α = .81).

Healthy diet. A total of 49 self-report items were administered to this sample (Goldberg & Strycker, 2002). Twenty of the items ask about specific health food practices (e.g., "When eating red meat, trim all visible fat?", "Have a vegetarian dinner?") and were rated on a 5-point Likert scale, ranging from (*usually or always*) to 5 (*N/A*). The other 29 items ask about the frequency of intake of various food items or liquids (e.g., French fried, Oat bran or what germ, 1% or skim milk), using a 5-point scale of 1 (*1< once/month*) to 5 *(≥ 5 times/week*). Following Goldberg and Strycker (2002), a total "healthy diet" composite was derived from all 49 items. Cronbach's alpha was .68 (ω = 89).

Exercise, substance use/smoking, and work avoidance. These criteria were derived from relevant items included in the "Behavioral Report Form" (Loehlin, 1976), as described in the relevant technical report by the Oregon Research Institute (Goldberg, 2008). The items describe past behaviour, for which the frequency is rated on a 5-point Likert scale from 1 (*never in my life*) to 5 *(≥ 15 times in past*). A total of 398 items were screened for construct-relevant criteria, resulting in three different clusters of items: exercise (five items), substance use/smoking (22 items), and work avoidance (five items), all of which involve motivation to either engage in adaptive behaviour or abstain from maladaptive behaviour and, to varying degrees, seem to require behavioural, cognitive, and emotional effort.

Each of these item clusters was submitted to a Principal Component Analysis to extract a common dimension, representative of these categories. All five exercise items (e.g., "Participated in an exercise program") loaded on a single dimension and indicated good reliability ($\alpha = .83$). The 22 substance use/smoking items (e.g., "smoked tobacco", "had a hangover") shared a common factor, but five items (aspirin or ibuprofen, antacids, tranquilising pills, laxative, no-doz or other stay-awake pills) did not load well ($\leq .30$) and were dropped, due to their distinct nature; items with loadings greater than .30 represented alcohol use, (hard) drug use, or smoking and had adequate reliability ($\alpha = .88$). The analysis for the five work avoidance items revealed two components. However, the second component was selected, because it represented the hypothetical construct: positive loadings of work avoidance behaviours ("was late for work" and "called in sick to work because I was too tired to get up") and negative loadings of work engagement behaviours ("stayed late at work", "went to work", and "stayed away from a social event in order to finish some work"); the first component had positive loadings of all five items, which is more indicative of work intensity. Internal reliability cannot be calculated for this component, because the corresponding items are bidimensional. McDonald's omega for this second component was .79 (Cronbach's alpha cannot be calculated for this component).

5.6.1.3. Statistical analyses

Bivariate correlations were used to examine the convergent, discriminant, and concurrent validities of the Drive:IPIP. A joint Varimax-rotated Principal Component Analysis with the 30 NEO-PI-R facets was conducted to examine the level of support for the hypothesised distinct factor within Five-Factor Model space. To examine incremental validity, hierarchical regression analysis was conducted for each criterion, controlling for higher-order personality dimensions at Step 1 and entering the Drive:IPIP total composite at Step 2. While for all criteria the NEO-PI-R domains were used as control variables, the primary and occupational scales of the HPI were used separately as predictors of work avoidance, given their occupational emphasis. Only the employed part of the sample (full- or part-time) was used for the analyses involving work avoidance; those who specified their job situation as homemaker, unemployed, retired, or “other” were excluded.

5.6.2. Results and discussion

5.6.2.1. Convergent and discriminant validity

Bivariate correlations involving both Drive forms and MEI scales are shown in Table 5.6. These correlations were consistently within a moderate range of .31 to .59 in the British sample and of .42 to .68 in the Norwegian sample. In both samples, correlations were slightly higher for the Drive:IPIP than for the Drive:S and highest with the MEI total score, as can be expected. The Drive:IPIP comprises multiple items for each facet and, thus, likely yields a somewhat more accurate measure of the GFM than the Drive:S. Regarding the MEI, the total score theoretically shares the largest amount of variance with the GFM, representing the construct more closely than its constituent facets. The magnitude of associations speaks to the convergent validity of the two Drive forms, but without indicating redundancy. The MEI represents a clinically slanted measure of current, clinically relevant motivational levels and energy (i.e., states), whereas the Drive:IPIP and Drive:S specifically measure the GFM and are based on a comprehensive sampling of motivation facets.

Intercorrelations among the Drive:IPIP, NEO-PI-R domains, and HPI scales are shown in Table 5.7. Predictions for associations were not made for all of the personality traits assessed with the NEO-PI-R domains and HPI scales. However, their associations are still included in the table, since all were involved in subsequent analyses of incremental validity. Consistent with the stated predictions, the Drive:IPIP did not correlate with Agreeableness, an interpersonal style that is conceptually distinct from motivation, whereas it correlated strongly with Conscientiousness. Conscientiousness was previously portrayed as comprising several motivational traits within the Five-Factor Model (Schmidt & Hunter, 1992), and a few of the Drive:IPIP facets (achievement-striving, self-discipline, diligence) derive from this domain. The observed moderate associations with Neuroticism and Extraversion are in accordance with previous findings of linkages between elements of trait motivation and these two personality dimensions (Zuckerman, Joireman, Kraft, & Kuhlman, 1999).

Concerning the HPI, the highest correlation was observed for Ambition, which is conceptually very similar to the behavioural factor derived from the Drive:IPIP facets. Moderate associations with Adjustment and Intellectance further speak to the convergent validity, with both entailing some degree motivation; Adjustment bears on the behavioural, cognitive factor, and affects factors in the derived structural model, whereas Intellectance resembles the cognitive and behavioural factors. The weak association with Prudence is not surprising, since although this variable was initially identified as a potential facet of the GFM, it was eventually removed from the representation due to poor fit. Also in line with expectations are the moderate associations observed for Clerical Potential, Sales Potential, and Managerial Potential, as all as the weak and non-significant associations with Service Orientation, School Success and Reliability, respectively. Overall, these results support the convergent and discriminant validity of the two Drive forms.

Table 5.7. *Study 5: Descriptive Statistics and Intercorrelations of Drive:IPIP, NEO-PI-R Domains, and HPI Primary and Occupational Scales in the Eugene-Springfield Community Sample*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Drive:IPIP	—																		
2. Neuroticism	-.57***	—																	
3. Extraversion	.56***	-.28***	—																
4. Openness	.26***	-.03	.38***	—															
5. Agreeableness	.03	-.22***	.02	.02	—														
6. Conscientiousness	.61***	-.46***	.18***	-.14**	.13**	—													
7. Adjustment	.37***	-.73***	.16***	.02	.31***	.23***	—												
8. Ambition	.69***	-.54***	.56***	.21***	-.12*	.40***	.45***	—											
9. Sociability	.29***	-.03	.62***	.42***	-.27***	-.07	.02	.45***	—										
10. Likability	.31***	-.27***	.44***	.21***	.48***	.09	.42***	.29***	.16***	—									
11. Prudence	.11*	-.26***	-.07	-.33***	.46***	.42***	.34***	.03	-.40***	.26***	—								
12. Intellectance	.34***	-.16***	.24***	.52***	-.19***	.07	.12**	.38***	.47***	.04	-.25***	—							
13. School Success	.22***	-.12**	.07	.23***	-.07	.11*	.11*	.22***	.12*	.01	.05	.32***	—						
14. Service Orientation	.19***	-.41***	.07	.04	.53***	.09	.63***	.14**	-.13**	.62***	.43***	-.03	.00	—					
15. Stress Tolerance	.50***	-.76***	.23***	.07	.13**	.30***	.89***	.58***	.10*	.31***	.22***	.21***	.16***	.46***	—				
16. Reliability	.06	-.38***	-.07	-.25***	.46***	.24***	.59***	.07	-.33***	.33***	.70***	-.20***	.02	.41***	.36***	—			
17. Clerical Potential	.55***	-.60***	.45***	.19***	.05	.30***	.64***	.76***	.31***	.38***	.17***	.32***	.22***	.31***	.72***	.25***	—		
18. Sales Potential	.46***	-.22***	.69***	.49***	-.17***	.03	.19***	.65***	.88***	.38***	-.35***	.58***	.18***	.03	.28***	-.27***	.48***	—	
19. Managerial Potential	.67***	-.55***	.46***	.14**	.01	.47***	.53***	.86***	.29***	.31***	.29***	.33***	.38***	.21***	.64***	.27***	.82***	.45***	—
N	496	475	475	475	475	475	476	476	476	476	476	476	476	476	476	476	476	476	476
M	3.72	80.09	105.21	113.27	125.57	124.90	24.15	20.76	10.55	18.33	19.70	13.32	8.33	9.50	18.17	11.57	17.06	36.60	27.74
SD	.43	23.71	20.24	21.63	16.91	19.36	6.87	5.72	4.86	3.21	4.10	4.41	3.20	2.47	4.83	3.45	3.82	9.30	5.19
Skewness	-.49	.51	-.29	-.12	-.52	-.42	-.48	-.76	.17	-1.44	-.45	-.15	-.44	-.48	-.74	-.41	-.29	-.18	-.67
Kurtosis	.89	.26	.07	-.07	.61	.86	-.37	.24	-.59	2.65	-.07	-.50	-.48	-.22	-.20	-.55	-.38	-.47	.61

Note. N = 458. Drive:IPIP = Drive: International Personality Item Pool version; NEO-PI-R = NEO Personality Inventory–Revised (Costa Jr. & McCrae, 1992); HPI = Hogan Personality Inventory (Hogan & Hogan, 1992).
*p < .05. **p < .01. ***p < .001.

Varimax-rotated Principal Component Analysis results are shown in Table 5.8. Whereas Eigenvalues and scree plot pointed to seven components, parallel analysis extracted only six (the scree plot and parallel analysis results are provided in Appendix 6). In line with theory (five personality domains plus one GFM), a six-factor solution was forced. The pattern of loadings clearly supports the Big Five plus one additional component, representing the GFM as a unique dimension. The cross-loadings of certain GFM facets on the Big Five domains substantiate the conceptualisation of the GFM as a construct that permeates personality-factor space. Conversely, there were no noteworthy loadings of the 30 NEO-PI-R facets on the GFM component; the highest was seen for E6: Positive Emotions at .31. Its emergence as a distinct dimension shows that the GFM extends beyond the boundaries of personality, or how people are like (i.e., behave, feel, and think), into the space of individual differences theorised to constitute motivation, the extent to which people pursue their goals and motives and execute other necessary tasks.

Table 5.8. *Study 5: Varimax-Rotated Component Matrix for Drive:IPIP and NEO-PI-R Facets in the Eugene-Springfield Community Sample*

	Factor loading					
Facet	C	N	O	A	E	GFM
Self-confidence	.31	.70				
Zest/enthusiasm/vitality	.32					.72
Valor/bravery/courage			.48			.35
Liveliness	.32					.59
Joyfulness		.58			.42	.39
Insight			.72			
Initiative	.71					
Diligence	.72					.39
Generates ideas		.33	.63			
Ind./persev./persis.	.73					.36
Self-discipline	.78					
Achievement-striving	.70					
Hope/optimism		.41				.59
N1		-.79				
N2		-.66		-.49		
N3		-.83				

N4		-.73				
N5	-.36	-.49		-.33		
N6	-.34	-.73				
E1					.78	
E2					.68	
E3	.35		.31	-.49	.36	
E4	.45				.37	
E5				-.46	.38	
E6					.69	.31
O1			.69			
O2			.70			
O3			.60		.43	
O4			.60			
O5			.74			
O6			.54			
A1		.46		.40	.45	
A2				.70		
A3				.59	.52	
A4				.75		
A5		-.39		.56		
A6				.59		

C1	.63	.52				
C2	.71					
C3	.67					
C4	.74					
C5	.79					
C6	.52			.33		
Eigenvalue	6.90	5.71	4.35	3.73	3.31	2.45
% of variance	16.05	13.27	10.12	8.68	7.69	5.71

Note. *N* = 475. Factor loadings of < .30 are omitted from the table. Drive:IPIP = Drive: International Personality Item Pool version; NEO-PI-R = NEO Personality Inventory–Revised (Costa Jr. & McCrae, 1992); GFM = General Factor of Motivation.

5.6.2.2. Concurrent and incremental validity

Bivariate correlations of the Drive:IPIP and Drive:S with motivation perceptions are displayed in Table 5.9. Correlations of the Drive:IPIP scores were within a moderate range of .62 to .69 for self-perceptions and weak-to-moderate for informant-perceptions of motivation (*r* = .26 and .38). The same correlations involving the Drive:S were somewhat stronger for self-perceptions (*r* = .69 to .77) and informant-perceptions (*r* = .36 and .53). These systematic differences between the two Drive forms reflect the measurement equivalence of the Drive:S and the criterion in the Online and British samples in terms of response scale and direct estimates. In contrast, the Drive:IPIP uses several specific items to measure each of the facets.

As shown in Table 5.9, Drive:S informant scores correlated moderately with participants' Drive:IPIP scores in the Online sample (*r* = .47) and with their Drive:S scores in both samples (*r* = .36 and .53); they correlated weakly with participants' Drive:IPIP scores in the British sample (.22). The correlations between Drive:S informant scores and participants self-perceptions of motivation was weak in the British sample (*r* = .22) and somewhat larger in the Online sample (*r* = 35). Drive:S informant scores correlated moderately to strongly with informant-perceptions of motivation (*r* = .78 and .57) at a level comparable to the associations between the same scores obtained from the participants.

Table 5.9. *Study 5: Descriptive Statistics and Bivariate Correlations of Drive:IPIP and Drive:S with Motivation Perceptions in the Online, British, and Norwegian Samples*

									Informant	
Variable	*N*	*M*	*SD*	Skewness	Kurtosis	Drive:IPIP	Drive:S	Perception	Drive:S	Perception
					Online sample					
Drive:IPIP	362	3.43	0.58	-0.33	-0.11	—				
Drive:S	351	61.10	16.63	-0.49	0.22	.81***	—			
Perception	351	61.50	25.22	-0.56	-0.41	.69***	.77***	—		
Drive:S (informant)	46	63.84	15.37	-0.17	-0.07	.47***	.53***	.35*	—	
Perception (informant)	46	63.91	22.66	-0.14	-0.81	.38**	.41**	.30*	.78***	—
					British sample					
Drive:IPIP	241	3.37	0.49	-0.01	-0.10	—				
Drive:S	233	61.28	14.36	-0.43	0.31	.73***	—			
Perception	233	63.18	21.35	-0.41	-0.38	.62***	.69***	—		
Drive:S (informant)	101	72.01	11.05	-0.23	-0.21	.22*	.36***	.28**	—	
Perception (informant)	101	76.28	19.57	-1.16	1.33	.26**	.37***	.39***	.57***	—
					Norwegian sample					
Drive:IPIP	142	3.59	0.49	-0.67	1.19	—				

Drive:S	142	4.94	0.88	-0.74	1.38	.83***	—	
Perception	142	5.26	1.34	-0.97	0.97	.67***	.71***	—

Note. Drive:S response scale differed for the Norwegian sample (see Method section). Drive:IPIP = Drive: International Personality Item Pool version; Drive:S = Drive: Short.
*p < .05. **p < .01. ***p < .001.

Table 5.10 shows the bivariate correlations of the Drive:IPIP, NEO-PI-R domains, and HPI scales with the five criteria derived from the Eugene-Springfield community sample data. These were significant and in the expected direction in all cases except for substance abuse, which did not correlate significantly with Drive:IPIP scores. In view of factors that can be expected to attenuate (small sample sizes involving informant-perceptions) or inflate the observed association (common-source and method variance), these result provide at least some preliminary support for the concurrent validity of the two Drive forms.

Table 5.10. *Study 5: Descriptive Statistics and Bivariate of Drive:IPIP, NEO-PI-R Domains, and HPI Primary and Occupational Scales in the Eugene-Springfield Community Sample*

Variable	Depression	Exercise	Healthy diet	Substance use	Work avoidance
Drive:IPIP	-.37**	.22**	.15**	.00	-.14*
Neuroticism	.46**	-.07	-.02	.03	.10
Extraversion	-.21**	.22**	.04	.12*	.09
Openness	.06	.11*	.21**	.22**	.17**
Agreeableness	-.03	-.05	.18**	-.32**	-.15*
Conscientiousness	-.19**	.05	.10	-.13**	-.22**
Adjustment	—	—	—	—	-.15*
Ambition					.00
Sociability					.15*
Likability					-.03
Prudence					-.21**
Intellectance					.05
School Success					-.04
Service Orientation					-.18**
Stress Tolerance					-.10
Reliability					-.18**
Clerical Potential					-.01
Sales Potential					.14*
Managerial Potential					-.10
N	469	477	398	454	256
Skewness	1.58	.20	-.01	.36	.18
Kurtosis	2.93	-1.15	-.43	.03	-.50

Note. Drive:IPIP = Drive: International Personality Item Pool version; NEO-PI-R = NEO Personality Inventory–Revised (Costa Jr. & McCrae, 1992); HPI = Hogan Personality Inventory (Hogan & Hogan, 1995).
*p < .05. **p < .001.

Incremental validity analyses for the Eugene-Springfield community sample criteria are shown in Table 5.11 (whole sample) and Table 5.12 (employed portion of the sample). The Drive:IPIP explained unique criterion variance beyond the Big Five personality traits in all criteria, including substance abuse. The criterion variance explained was not particularly large for any of the predictors, indicating that other variables (external factors) than psychological constructs and individual differences possibly carry more weight. The significant explanatory effect of the Drive:IPIP on substance abuse suggests that motivation has desirable and undesirable effects on substance abuse that cancel each other out to an overall non-significant bivariate correlation. As a result of "trying hard" in life, people who are very motivated may need to cope with the ensuing stress and, consequently, may resort to maladaptive coping strategies, such as drinking and smoking. On the other hand, motivation may also lead people to abstain from substance use, as indicated by the significant predictive effect of motivation when controlling for the Big Five domains; the Big Five seem to partial out the undesirable effects of motivation.

Concerning work avoidance, the results show that the Drive:IPIP predicts unique criterion variance irrespective of whether the NEO-PI-R domains, HPI primary scales, or HPI occupational scales are controlled for as predictors. Again, the variance was not very large for any measure, but the incremental variance explained by the Drive:IPIP went up to 6.1% (for the NEO-PI-R) and the Drive:IPIP also showed the largest beta weight when included with the NEO-PI-R domains or HPI primary scales.

Keeping common-source and method variance for some of the reported analyses in mind, these result provide good preliminary support for the concurrent, predictive, and incremental validity of the Drive:IPIP.

Table 5.11. *Study 5: Hierarchical Regression Analyses Predicting Criteria with the NEO-PI-R Domains (Step 1) and the Drive:IPIP (Step 2) in the Eugene-Springfield Community Sample*

	Depression			Exercise			Healthy diet			Substance use		
Step 1	$F(5,444) = 28.06^{**}$, $\Delta R^2 = .240^{**}$, $R^2_{Adj} = .232$			$F(5,451) = 5.04^{**}$, $\Delta R^2 = .053^{**}$, $R^2_{Adj} = .042$			$F(5,377) = 7.39^{**}$, $\Delta R^2 = .089^{**}$, $R^2_{Adj} = .077$			$F(5,429) = 17.05^{**}$, $\Delta R^2 = .166^{**}$, $R^2_{Adj} = .156$		
Step 2	$F(6,443) = 28.18^{**}$, $\Delta R^2 = .036^{**}$, $R^2_{Adj} = .266$			$F(6,450) = 6.20^{**}$, $\Delta R^2 = .023^{**}$, $R^2_{Adj} = .064$			$F(6,376) = 7.09^{**}$, $\Delta R^2 = .012^{*}$, $R^2_{Adj} = .087$			$F(6,428) = 15.27^{**}$, $\Delta R^2 = .011^{*}$, $R^2_{Adj} = .165$		
Step 2 predictors	β	Tolerance	VIF	β	Tolerance	VIF	β	Tolerance	VIF	β	Tolerance	VIF
N	.37**	.62	1.62	.05	.61	1.63	.13*	.60	1.65	-.12*	.61	1.65
E	-.03	.63	1.59	.11*	.62	1.61	-.12	.63	1.60	.12*	.62	1.62
O	.18**	.74	1.35	-.02	.74	1.35	.19**	.72	1.39	.23**	.73	1.37
A	.04	.91	1.10	-.03	.92	1.09	.19**	.89	1.12	-.35**	.92	1.08
C	.21**	.50	1.98	-.11	.50	2.01	.05	.49	2.04	-.02	.50	1.99
Drive:IPIP	-.33**	.33	3.07	.27**	.32	3.13	.20*	.31	3.26	-.18*	.32	3.16

Note. NEO-PI-R = NEO Personality Inventory–Revised (Costa Jr. & McCrae, 1992); Drive:IPIP = Drive: International Personality Item Pool version; VIF = variance inflation factor, N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness.
**p* < .05. ***p* < .001.

Table 5.12. *Study 5: Hierarchical Regression Analysis Predicting Work Avoidance with the NEO-PI-R Domains (Step 1a), HPI Primary Scales (Step 1b), or HPI Occupational Scales (Step 1c) and the Drive:IPIP (Steps 2a, 2b, and 2c) in the Eugene-Springfield Community Sample*

Step 1a	$F(5,250) = 4.60^{***}$, $\Delta R^2 = .084^{***}$, $R^2_{\text{Adj}} = .066$
Step 1b	$F(7,248) = 2.30^{*}$, $\Delta R^2 = .061^{*}$, $R^2_{\text{Adj}} = .035$
Step 1c	$F(6,249) = 4.14^{***}$, $\Delta R^2 = .091^{***}$, $R^2_{\text{Adj}} = .069$
Step 2a	$F(6,249) = 4.89^{***}$, $\Delta R^2 = .021^{*}$, $R^2_{\text{Adj}} = .084$
Step 2b	$F(8,247) = 3.05^{**}$, $\Delta R^2 = .029^{**}$, $R^2_{\text{Adj}} = .060$
Step 2c	$F(7,248) = 4.22^{***}$, $\Delta R^2 = .016^{*}$, $R^2_{\text{Adj}} = .081$

Step 2 predictors	β	Tolerance	VIF
(Step 2a)			
Neuroticism	-.06	.65	1.53
Extraversion	.14	.70	1.42
Openness	.16*	.75	1.34
Agreeableness	-.13*	.91	1.10
Conscientiousness	-.08	.56	1.79
Drive:IPIP	-.23*	.38	2.60
(Step 2b)			
Adjustment	-.10	.64	1.57
Ambition	.13	.43	2.34
Sociability	.07	.54	1.86
Likability	.07	.74	1.35
Prudence	-.15	.64	1.56
Intellectance	.02	.67	1.49
School Success	-.02	.85	1.18
Drive:IPIP	-.23**	.54	1.86
(Step 2c)			
Service Orientation	-.14	.66	1.52
Stress Tolerance	-.01	.40	2.48
Reliability	-.05	.58	1.71
Clerical Potential	.22	.22	4.47

Sales Potential	.19*	.52	1.92
Managerial Potential	-.21	.24	4.22
Drive:IPIP	-.17*	.56	1.80

Note. NEO-PI-R = NEO Personality Inventory–Revised (Costa Jr. & McCrae, 1992); HPI = Hogan Personality Inventory (Hogan & Hogan, 1995); Drive:IPIP = Drive: International Personality Item Pool version; VIF = variance inflation factor.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

5.7. General Discussion

The results fully support the validity of the GFM as operationalised in the present investigation. In particular, they support the existence of a single second-order factor, which was found to explain a considerable part of the variance in the motivation facets derived from the IPIP (in excess of 70%). This result is in line with the properties of similar measures of motivation or cognate constructs, which assume a unidimensional structure of factors that load on a global composite (Duckworth et al., 2007; Fehnel et al., 2004). Presumably, these measures share considerable variance with, and represent parts of, the GFM. As has been previously demonstrated for other constructs, such as trait EI (Petrides, Pita, et al., 2007), the results also demonstrate that facet-level traits linked to different personality dimensions share variance not encompassed by the established personality taxonomies, such as the Big Five domains. A systematic source of variance of selected facet traits appears to be the proposed GFM.

Although the GFM naturally permeates, and in fact overlaps with general personality domains, especially as operationalised here, it is far from redundant with them. For example, the maximum correlation with the Big Five personality traits was between the Drive:IPIP and Conscientiousness, with which it shares a few facets, at r = 61. In comparison, a correlation of .77 was observed between the similar construct of

Grit and Conscientiousness, and the Grit Scale still evidenced incremental validity vis-à-vis the Big Five domains (Duckworth et al., 2007; Duckworth & Quinn, 2009). Likewise, the GFM explained incremental variance in all criteria examined in the present study when controlling for the Big Five and, in the case of work avoidance, also for the HPI primary and occupational scales. These results strongly support the distinctiveness and utility of the GFM, relative to established personality dimensions. Accordingly, they build on previous accounts that the Five-Factor Model does not yield a suitable taxonomy of motivational traits (Kanfer & Ackerman, 2000), which appear to be distributed across conceptually orthogonal personality dimensions (if unevenly) and possibly go beyond them.

5.7.1. Contributions

To a certain degree, the field of motivation (from an individual differences and psychometric perspective) has been stepping in the dark at a conceptual level in its virtually exclusive focus on motives (Mayer et al., 2007). While motives constitute a valid and useful set of attributes, they must not be confounded with motivation, as argued in the introduction. Motives were found to be multidimensional, involving multiple constructs (Bernard, 2009; Bernard, Mills, Swenson, & Walsh, 2008; Jackson, 1984; Reiss, 2004; Reiss & Havercamp, 1998), whereas motivation is single, unidimensional construct, as the present results demonstrate. A main contribution of the present work lies in the explicit distinction of motivation from motives, one that has been largely omitted in the scientific literature.

Even in the unlikely scenario that a model or measure exists that represents the GFM concisely, it would not have been developed intentionally or systematically for this purpose. The GFM has conceptual appeal over cognate, novel-sounding constructs.

For example, everyone knows what motivation is, whereas a large number of people will not be familiar with the concept of “Grit”. Motivation is a concept that, for good reason, has been part of our common language for quite some time. Yet, it may have been somewhat misconstrued as regards conceptual confounding with motives.

Another major contribution is that a GFM representation and structural model was systematically derived and validated. Surprisingly few measures can be found that may tap into the GFM to some degree. These are the MEI (Fehnel et al., 2004), the Motivation and Engagement Scale (Liem & Martin, 2012), and possibly also the Grit Scale (Duckworth et al., 2007; Duckworth & Quinn, 2009). However, these measures have limited and uncertain construct validity as far as the GFM is concerned, since none of them has been developed to measure the construct specifically. Moreover, the MEI has a clinical emphasis and assesses state motivation, focusing on current motivation levels. The GFM model advanced here was not only developed to represent the GFM explicitly, but it also derives from a comprehensive set of human attributes, increasing the likelihood of full construct coverage. Overall, the contributions made by the present research span the conceptualisation, operationalisation, and validation, of the proposed GFM.

5.7.2. Future directions

At present, only a single structural model with two measures exists for the proposed construct. For the field to (re)gain momentum, it will be necessary that alternative measures, based on unique, yet solid approaches to sampling from the construct domain, be devised. On the one hand, a plethora of models and corresponding measures often complicates the integration of research findings, particularly where measures vary considerably in scope and focus. Still, convincing and converging results from multiple measures of the GFM would speak to the value of the construct and the research that surrounds it. The key point to be considered is that multiple GFM models and measures ought to be grounded in the same general definition of the construct (e.g., one that does not overlap motives and cognate types of constructs).

Future psychometric efforts will need to further investigate the nomological network and criterion validity of the GFM, using other relevant constructs not considered in the present investigation. Although the construct was defined in detail here, its operational vehicle(s) should undergo systematic (convergent and discriminant) validation efforts, ideally within a multitrait-multimethod framework (e.g., Ziegler, 2014b). For example, the distinctiveness between the GFM and motives (and cognate constructs) remains to be examined. If successful, the focus will naturally shift to the implications of the construct in the real world, by examining predictive effects on relevant, objectively assessed real-life criteria, such as academic achievement and work performance. The Drive, which can be viewed as an overarching label for a family of measures, will provide a solid operational vehicle for these endeavours. The results gathered here, especially those involving informant ratings, are very promising in this respect.

5.8. Conclusions

The present chapter described the development and validation of a measure of the GFM, grounded in a conceptual distinction between motivation and motives. A systematically derived set of facets, operationalised via the IPIP (Goldberg et al., 2006), was used to develop a clean and replicable structural model of the GFM. The model comprised three first-order factors representing affective, behavioural, and cognitive aspects of motivation. Moreover, the model was further confirmed using a short form (the Drive:S), which converged highly with the Drive:IPIP. The global composite, representing the GFM, demonstrated convergent and discriminant validity with conceptually similar and distinct personality constructs, as well as concurrent and incremental validity of relevant criteria. However, some of the derived facets may be redundant or extraneous, which is why it would be advantageous to examine FB in the context of this new operationalisation. That was the aim of **Chapter 6**.

CHAPTER 6: Application of Facet Benchmarking in the Context of the General Factor of Motivation

6.1. Introduction

In its first investigation (**Chapter 2**), FB was applied to a 15-facet measure of trait EI, using existing data collected in validation studies of the measure. The criteria used to derive the criterion-based composites varied across samples in number and types. Despite expected variability in the explanatory effects of the facets across samples, four of the 15 samples did not explain unique variance in the criterion-based composite in any sample. An additional fifth facet only had a significant (i.e., in a direction opposite to that predicted by the other facets), yet unexpected beta weight in some of the samples. Compared to the original 15-facet composite, a modified composite of the 10 remaining facets showed consistently higher associations with the criterion-based composite, and the five problem facets were identified as redundant. A limitation of this preliminary application of FB was that, although a wide range of criteria were administered across the six samples, the variance of some trait EI facets may still have been underrepresented in the various criterion-based composites. The criteria were neither systematically nor deliberately selected for the purpose of applying FB.

The second investigation used data specifically collected for the purpose of scrutinising FB, this time in the context of dispositional mindfulness (**Chapter 4**). In order to derive a comprehensive, criterion-based representation of the construct variance, the criteria were systematically selected, based on their usage in validation studies of the existing mindfulness scales. This set of criteria was administered, along with multi-faceted mindfulness scales of up to five facets, to three samples, and the results were cross-validated on a fourth sample, using a completely different set of criteria. Across samples, the same facet was identified as not occupying unique variance in the criterion-based composites and as weakening the measure's associations with the criterion-based composite. Subsequently, this facet was identified as extraneous, corroborating emerging findings of other problematic psychometric attributes.

An open question concerns the mono-method assessment of facets and criteria in the applications of FB conducted so far. Using exclusively self-report measures, especially of the same scale format, may introduce method effects, which could differentially influence the associations between and criteria. Moreover, although the application to mindfulness did much to address this concern, it is necessary to utilise a broad set of criteria, representative of all facets; each facet should be represented in the variance of the initial set of criteria, even though individual criteria may be discarded at Step 2 of FB.

The purpose of this study was threefold. First, it was aimed at further increasing confidence in the method's efficacy, by demonstrating it with reference to a new construct. To this end, FB was applied to the GFM, as conceptualised and operationalised in **Chapter 5**. The second objective was to assess if common-method effects may bias the results, by introducing any confounding that would need to be taken into consideration. Thus, for some of the analyses executed, the variables used to derive the criterion-based composites and facets differed in response scale format and, in some instances, even in the source of the ratings provided. A third way in which this investigation sought to establish the integrity of FB was to maximise certainty that its efficacy is not simply explained by an underrepresentation of the hypothetical facets in the pool of variables used to extract the criterion-based composites. To address this issue, part of this investigation used a criterion-based composite theoretically representative of all hypothetical facets; for each facet, a conceptually similar criterion was selected, and the criterion-based composite was then derived from the aggregate of these criteria. Subsequently, different criteria were used to increase certainty in the existence of any problem facets.

Given the rigorous design of the study, it was suitable for comparing principal component extraction, the standard method for deriving the criterion-based composite in FB, to principal axis factoring, which some may regard as a conceivable alternative. Rotation remains wholly inappropriate for this purpose, since the focus is to ascertain loadings of theoretically criteria on the theoretical factor, not to derive latent factors and thereby impose a simple structure that disguises more accurate loadings.

6.2. General Method

The application of FB in this study was spread out across three parts, involving a total of four different samples. In all samples, the Drive:IPIP was used as the operational vehicle of the GFM facets. Parts I and II both encompass Steps 1 to 3 of FB, with Part II constituting a cross-validation on different types of criteria. Part III presents Steps 4 and 5, which were contingent on the pattern of results from Parts I and II.

6.2.1. Samples

The samples in this investigation were those already used in **Chapter 5**. However, the effective sample size different for three of the samples. In one case, the data collection continued, whereas in two cases, the use of additional or fewer variables meant that a larger or small number of cases could be used, respectively. Accordingly, amended sample descriptions are provided below.

Online sample

This sample (*N* = 351; 78.3% female) was recruited using recruitment platforms for academic research in psychology (e.g., onlinepsychresearch.co.uk, callforparticipants.com; *n* = 43.3%), a commercial recruitment platform for online academic research (findparticipants.com; *n* = 34.5%), and “tweets” by relevant authorities or platforms (*n* = 22.2%). The mean age of the sample was 33.3 years (*SD* = 13.4, range = 15.9–73.1). Ethnic backgrounds of the participants were 76.9% Caucasian, 6.6% Asian, 4.6% African, 3.7% South Asian (Bangladesh, India, and Pakistan), and 8.3% other, unspecified, or mixed. Educational qualifications obtained were distributed across GCSE/O or similar (6.0%), A Level or similar (19.7%), BA/BSc or similar (34.2%), MA/MSc or similar (20.3%), MBA (3.4%), PhD (2.3%), and other (14.0%). Most of the participants were studying (33.9%) or working (43.6%) full-time; smaller proportions were studying (9.1%) or working (18.2%) part-time. Participants who provided an email address were entered into a price draw for one of several gift vouchers.

British sample

This sample (*N* = 233; 80.3% female) was recruited using the participant pool of a major British university. Most participants (88.2%) were full-time students at the university and pursuing an undergraduate (69.6%) or Master’s (25.4%) degree. Small proportions were working on a PhD (2.1%), studying part-time (2.9%), and/or working full- (6.7%) or part-time (8.0%). Participant ages had a mean of 22.3 years (*SD* = 22.3) and ranged from 18.0 to 75.1 years. Most participants were of Caucasian (50.4%) and Asian (38.7%) ethnic backgrounds; the remainder came from South Asian (Pakistan, India, Sri Lanka; 3.4%), African (0.8%), and other, unspecified or mixed backgrounds (6.7%). Most participants were compensated with course credit, and all were entered into a prize draw for gift vouchers as a token of appreciation.

Norwegian sample

This sample (*N* = 143, 61.3% female) was recruited from Norway, using Qualtrics Sample Finder. Its mean age was 47.6 years (*SD* = 7.7) and a large proportion (96.5%) was of Caucasian descent. While it was expected that several participants were not native English speakers, English is an established second language in Norway. This sample reported an average of 33.8 years (*SD* = 9.5) of spoken English. Highest educational qualifications were High school or similar (24.6%), BA/BSc or similar (22.5%), A Level, IB, or similar (19.7%), MA/MSc or similar (14.1%), MBA (9.9%), PhD (4.9%), and other (4.2%). A minority of the sample (9.9%) was still enrolled in part- or full-time education; approximately half the sample (47.9%) were full-time and 21.1% part-time workers. This sample was financially compensated.

Eugene-Springfield community sample

The data for this widely used U.S. sample was provided by the Oregon Research Institute. Of the original dataset, 208 participants (54.8% female) had complete data on all key variables in the present investigation. This portion of the sample had a mean age of 45.1 years (*SD* = 8.4, range: 21 to 69 years) and most of them were Caucasian adults (97.6%). The remaining ethnicities were Asian American, Hispanic, or Native American. The most common educational backgrounds were “some college” (29.3%), "post-college degree" (24.0%), "college graduate" (24.0%), and "some post college" (12.0%); the remaining were "vocational/technical schooling" (5.3%), "high school graduate" (4.8%) and "not graduated from high school" (0.5%). Participants worked full-time (75.5%) or part-time (24.5%).

6.2.2. Measure and procedure

The development of the Drive:IPIP described in **Chapter 5** resulted in 13 facets that loaded on three factors, provisionally labelled Affective, Behavioural, and Cognitive Motivation. The items were responded to on the standard IPIP Likert scale, ranging from 1 (*very inaccurate*) to 5 (*very accurate*). Table 6.1 shows the three factors and their constituent facets, along with sample items and internal reliability coefficients for each sample. Cronbach's alphas for the facet scores in the four study samples, in the order presented above, were .94, .92, .93, and .88.

Table 6.1. *Operationalisation of the General Factor of Motivation via the Drive:IPIP: Factors, Facets, Sample Items, and Internal Reliabilities*

		Cronbach's α			
Factors and facets	Sample item	Online sample (N = 351)	British sample (N = 233)	Norwegian sample (N = 143)	ESCS (N = 208)
Affective motivation					
Joyfulness	Radiate joy.	.85	.79	.84	.74
Liveliness	Tire out quickly.*	.85	.83	.85	.76
Hope/optimism	Will succeed with the goals I set for myself.	.81	.70	.78	.69 (.80)
Zest/enthusiasm/vitality	Don't approach things half-heartedly.	.82	.78	.78	.76
Self-confidence	Am sure of my ground.	.78	.70	.56 (.75)	.72
Behavioural motivation					
Ind./persev./persis.	Don't quit a task before it is finished.	.81	.77	.75	.79
Diligence	Stop when work becomes too difficult.*	.70	.70	.63 (.77)	.62 (.77)
Self-discipline	Carry out my plans.	.78	.81	.54 (.73)	.70
Initiative	Get chores done right away.	.81	.75	.75	.75
Achievement-striving	Go straight for the goal.	.72	.81	.66 (.79)	.62 (.76)
Cognitive motivation					
Generates ideas	Quickly think up new ideas.	.85	.75	.76	.79
Insight	Put a new perspective on things.	.76	.78	.69 (.81)	.75

Valor/bravery/courage	Avoid dealing with awkward situations.*	.81	.84	.72	.78

Note. Reverse-scored sample items are denoted by an asterisk. Where Cronbach's alpha is low (< .70), McDonald's omega is given in parentheses. IPIP = International Personality Item Pool (Goldberg et al., 2006); ESCS = Eugene-Springfield community sample; Ind./persev./persis. = Industriousness/perseverance/persistence.

The Online, British, and Norwegian samples participated via an electronic survey. After providing demographic information, they completed the Drive:IPIP and measures of the criteria used and described in either Part I or Part II. Since the items were administered electronically, their order of appearance was randomised across participants. Also, any missing responses were automatically highlighted while completing the survey. The Eugene-Springfield community sample completed the measures by hand, as administered by the Oregon Research Institute.

6.3. Part I

Part I aims to maximise confidence in FB, by ascertaining that its hitherto very promising results are not a function of (a) some facets being unrepresented in the variables used to extract the criterion-based composites and (b) method effects. Objective A was approached by including a criterion uniquely relevant for each facet and known share variance with the GFM. Specifically, direct estimates on each of the facets were used to derive an alternative representation of the variance. In order to examine the impact of any method effects (Objective b), two of the samples completed the estimates on a different type of response scale than that used for the Drive:IPIP items, whereas the third sample gave these estimates on the same type of scale. Furthermore, informant ratings were obtained from two of the samples to assess the impact of common-source bias. Also, principal axis factoring was used in order to examine the results of an alternative extraction method for deriving the criterion-based composite. The relevant data were collected from the Online, British, and Norwegian samples.

6.3.1. Method

6.3.1.1. Informants

Participants in the Online and British samples were asked to nominate an informant. To achieve an adequate sample size, these data were combined for these two samples, resulting in a total of 147 informants (68.7% female), aged between 16.2 to 69.2 years (M = 31.7, SD = 14.4). The informants had known the participants who nominated them for an average of 12.9 years (SD = 10.5, range = 0.1 to 46.4) as parents (23.1%), siblings (16.3%), friends (40.1%), spouses/romantic partners (18.4%), or in a different capacity (2.0%).

6.3.1.2. Measures of criteria

Online, British, and Norwegian samples: Facet self-estimates

Participants gave direct estimates of each of the Drive facets, which provided the basis for the Drive:S in **Chapter 5**. The Online and British samples completed the items on a visual analogue scale, ranging from 0% to 100%. Using an electronic slider, participants were prompted to rate themselves in comparison to other people of similar age. The Norwegian sample completed the items on a 7-point Likert scale, ranging from 1 (*very little*) to 7 (*very much*). Administration instructions for these facet ratings were similar to the Drive:IPIP, but they were modified to accommodate the respective scale and response format.

Online and British samples: Facet informant-estimates

The informants nominated by the Online and British samples also provided estimates on each of the facets. They used the same visual analogue scale (ranging from 0% to 100%) to rate the participants.

6.3.2. *Results and discussion*

Step 2 of FB involved the extraction of the first principal component from the facet estimates. This component was subsequently regressed on the Drive:IPIP facets, using the stepwise algorithm and starting with all facets at the initial step (Step 3 of FB) to distinguish between predictive and non-predictive facets.

6.3.2.1. Dimensional reduction of criteria

Principal Component Analysis results for the facet estimates in each sample are presented in Table 6.2. Across samples, all facet estimates loaded on the first principal component; no variables needed to be considered for exclusion from the criterion-based composite. In theory, the criterion-based composite was, thus, representative of all Drive facets in this part of the study.

When principal axis factoring was used, all facet estimates continued to load well on the first factor and were comparable to principal component loadings. In fact, the correlations between principal components and principal axis factors across the four analyses reported in Table 6.2 were .10, .10., .99, and .10. Factor analysis (principal axis factoring) results are not reported for the sake of brevity, but they are available upon request.

Table 6.2. *Part I: First Principal Component Loadings for Facet Estimates*

Criteria	Online sample (N = 351)		British sample (N = 233)		Informants (N = 147)		Norwegian sample (N = 143)	
	Factor loading	Communality	Factor loading	Communality	Factor loading	Communality	Factor loading	Communality
Self-confidence	.74	.58	.67	.53	.52	.43	.65	.46
Zest/enthusiasm/vitality	.81	.80	.80	.80	.79	.74	.77	.69
Valor/bravery/courage	.68	.49	.63	.53	.62	.59	.72	.52
Liveliness	.76	.73	.78	.82	.74	.72	.74	.67
Joyfulness	.73	.84	.72	.82	.76	.80	.74	.78
Insight	.56	.69	.57	.65	.43	.53	.60	.45
Initiative	.70	.60	.65	.56	.77	.64	.77	.60
Diligence	.65	.72	.57	.73	.56	.74	.57	.49
Generates ideas	.63	.75	.63	.69	.50	.69	.74	.56
Ind./persev./persis.	.69	.68	.57	.62	.61	.82	.55	.57
Self-discipline	.53	.58	.55	.58	.53	.74	.51	.39
Achievement-striving	.65	.61	.62	.69	.59	.55	.57	.41
Hope/optimism	.75	.74	.74	.73	.74	.71	.67	.65
% of variance	47.35		43.61		40.57		44.70	

Note. Ind./persev./persis. = Industriousness/perseverance/persistence.

6.3.2.2. Regression of criterion-based composite on Drive:IPIP facets

Stepwise regression summaries for the criterion-based composite regressed on the Drive:IPIP facets are shown in Table 6.3. In the interest of space, results for the initial and final models, including only beta weights for facets included in the final model of each regression analysis, are presented. Two facets, self-discipline and achievement-striving, did not make it into the final model in any of the four samples. Consequently, they are omitted from Table 6.3. All beta weights of facets included in the final models had the expected positive sign and, thus, none of them needed to be removed manually.

Regression of the criterion-based composite derived from the informant-estimates on the facets yielded only one predictive facet (joyfulness). Although different from the other analyses, the important observation is that this analysis did not reveal additional significant facets that were not already identified by the other analyses. These results provide no evidence that response bias suppressed the effects of any valid facets in the analyses of the data based on self-report only; possibly inflated associations between some facets and the criterion-based composite did not seem to impinge on the associations of other valid facets less prone to response bias. On the other hand, it seemed difficult for any facet to account for unique variance in these informant-based construct representations.

Where the nearly identical principal axis factors were regressed on the 13 facets, the same pattern of significant predictors emerged in three of the four analyses shown in Table 6.3. Only a minor difference was seen in the Online sample, where one facet less was significant when using the principal axis factor; the beta weight of Valor/bravery/courage lost its statistical significance. These results suggest that the two extraction method yield largely equivalent results at Step 3 of FB, but also that principal component extraction may be less likely to drop facets, which reflects the way in which it extracts variables (i.e., incrementally). Again, these results will not be displayed here, but can be requested from the author.

Table 6.3. *Part I: Stepwise Regression Analysis Summaries for Drive:IPIP Facets Predicting the Criterion-Based Composite*

	Online sample (*N* = 351)			British sample (*N* = 233)			Informants (*N* = 147)			Norwegian sample (*N* = 143)		
Drive:IPIP facets	β	*F*	R^2_{Adj}	β	*F*	R^2_{Adj}	β	*F*	R^2_{Adj}	β	*F*	R^2_{Adj}
Model 1 (all facets)		55.10***	.67		22.54***	.55		2.46**	.11		25.99***	.70
Final model		90.28***	.67		71.57***	.55		22.63***	.13		54.62***	.69
Self-confidence	.16***											
Zest/enthusiasm/vitality										.16*		
Valor/bravery/courage	.09*			.17***						.17**		
Liveliness	.13**			.29***						.15*		
Insight	.12*									.26***		
Initiative	.13*											
Diligence	.11*			.26***								
Generates ideas	.13**											
Ind./persev./persis.										.19**		
Hope/optimism	.22***									.19*		
Joyfulness				.27***			.37***					

Note. Only beta weights for facets retained in the final models are displayed. IPIP = International Personality Item Pool; Ind./persev./persis. = Industriousness/perseverance/persistence.

6.4. Part II

Part II was aimed at cross-validating the results, using a different approach to capturing the construct variance in the criterion-based composite. Despite constituting proxy representations of the GFM, the criterion-based composites used in this part of the study were unlikely to give as accurate a representation of the facets as the estimates specifically aligned with each facet in Part I. Therefore, no additional facets should emerge as predictive of the criterion-based composites (measurement error aside); fewer facets should be predictive in this instance, because the construct domain is less likely to be represented comprehensively in this part of the study.

In the British and Norwegian samples, the criterion-based composite was derived from a state measure and global perceptions of motivation. Although neither of these two measures may represent the construct with 100% accuracy, both are intended to tap into the GFM at the global construct level and should give a more accurate representation of the construct when combined. Given unique limitations associated with each measure, their integration should align the criterion-based composite more closely with the GFM than each measure alone. In the Eugene-Springfield community sample, the criterion-based composite was derived from five specific criteria associated with the GFM, as used in **Chapter 5**: depression, exercise, healthy diet, substance use, and work avoidance.

6.4.1. Method

6.4.1.1. Measures of criteria

British and Norwegian samples

State motivation. Participants (n = 171) completed the MEI (Fehnel et al., 2004), which consists of three interrelated factors, for which a common factor was not previously presented: Mental Energy (10 items), Social Motivation (10 items), and Physical Energy (7 items). The MEI does not use a uniform question format and corresponding response scale across its 27 items. Most of the items are measured on a 6- or 7-point Likert scale (frequency type), ranging from either *Never* to *Every day or nearly every day* (e.g., “During the past 4 weeks, how often did you feel enthusiastic when you began your day?”), *Never* or *None of the time* to *All of the time* (e.g., “During the past 4 weeks, how often did you avoid social conversations with others?”), or *Never* to *At least 7 times a week* (e.g., During the past 4 weeks, how often did you engage in recreational activities or hobbies?). Another six items are “to what extent” questions based on a 5-point Likert scale, ranging from *Not at all interested* to *Extremely interested* (e.g., “During the past 4 weeks, to what extent were you interested in learning or trying new things?”). Cronbach’s alphas for the three factors were .84 (.89) for Mental Energy, .83 (.81) for Social Motivation, .85 (.85) for Physical Energy (alphas for the Norwegian sample are reported in parentheses).

General motivation estimate. Participants provided a general motivation estimate, along with their facet estimates. As described and explained in Part I, participants in the British sample provided this estimate on a visual analogue scale (possible range = 0–100%), while those in the Norwegian sample responded on a 7-point Likert scale.

Eugene-Springfield community sample

Depression. A modified version of the Center for Epidemiologic Studies Depression Scale (Radloff, 1977) was used. The scale, which assesses depressive symptomatology in non-clinical populations, was extended by four items (Goldberg, 2008). A total of 24 depressive symptoms during the past week (e.g., “I had a poor appetite”) were responded to on a 5-point Likert scale, ranging from 1 (*not at all past week*) to 5 (*most or all of the time*). Cronbach’s alpha was .79.

Healthy diet. Following Goldberg and Strycker (2002), a total “healthy diet” composite reflecting food consumption behaviour was measured using 49 self-report items. Twenty of the items concern specific health food practices (e.g., “When eating red meat, trim all visible fat?”, “Have a vegetarian dinner?”) and are rated on a 5-point Likert scale, ranging from 1 (*usually or always*) to 5 (*N/A* = not applicable). The other 29 items concern the consumption frequency of various food items or liquids (e.g., French fried, Oat bran or wheat germ, 1% or skim milk), indicated on a 5-point scale of 1 (*1< once/month*) to 5 *(≥ 5 times/week*). Cronbach’s alpha was .70.

Exercise, substance use/smoking, and work avoidance. These three criteria were extracted separately from selected items in the Behavioral Report Form (Goldberg, 2008; Loehlin & Nichols, 1976) in **Chapter 5**. For each criterion, the frequency of relevant past behaviours was rated on a 5-point Likert scale from 1 (*never in my life*) to 5 *(≥ 15 times in past*). Numbers of items per criterion and their internal consistency in this sample were five for exercise (Cronbach’s α = .84), 22 for substance use/smoking (Cronbach’s α = .88), and five for work avoidance (McDonald’s ω = .79).

6.4.1.2. Statistical analyses

The MEI factors were treated as indicators of a latent construct (state motivation), which was combined with the general motivation estimates into a global composite via Principal Component Analysis. A second Principal Component Analysis was conducted on the five criteria in the Eugene-Springfield community sample. The criterion-based composites derived at this step (Step 2 of FB) were subsequently regressed on the Drive:IPIP facets (Step 3 of FB), using stepwise regression with all 13 facets in the initial regression model.

6.4.2. Results and discussion

6.4.2.1. Dimensional reduction of criteria

Using Principal Component Analysis, the criterion-based composite was derived from state motivation, as measured with the MEI, and general motivation estimates in the British and Norwegian samples. Beforehand, it was necessary to ascertain that a general factor explains the shared variance in the MEI scales, which were, thus, submitted to independent Principal Component Analyses in each sample. These results are shown in Table 6.4. In both samples, all three MEI scales loaded strongly on a single component that explained much of the scales' shared variance. The derived composite was then combined with general motivation estimates. Correlations between these two variables were moderate at .48 in the British sample and .60 in the Norwegian sample, which speaks to shared variance of these variables, reflective of the GFM.

Table 6.4. *Part I: Principal Component Loadings for MEI Scales in the British and Norwegian Samples*

	British sample (N = 171)		Norwegian sample (N = 143)	
MEI scales	Factor loading	Communality	Factor loading	Communality
Physical energy	.89	.80	.90	.81
Social motivation	.74	.54	.86	.74
Mental energy	.81	.65	.89	.79
% of variance	66.31		77.88	

Note. MEI = Motivation and Energy Inventory (Fehnel et al., 2004).

Principal Component Analysis results for the criteria in the Eugene-Springfield community sample are shown in Table 6.5. The five criteria all loaded satisfactorily on the first principal component. Therefore, the criterion-based composite was derived from all criteria available in this sample.

Table 6.5. *Part II: First Principal Component Loadings for Criteria in the Eugene-Springfield Community Sample (*N *= 208)*

Criteria	Factor loading	Communality	% of variance
Healthy diet	-.74	.70	24.84
Work avoidance	.33	.57	
Depression	.50	.79	
Exercise	-.40	.77	
Substance use/smoking	.42	.61	

6.4.2.2. Regression of criterion-based composites on facets

Table 6.6 shows the summaries of the regressions of the criterion-based composites derived from either state motivation and general motivation estimates (British and Norwegian samples) or the five construct-relevant criteria (Eugene-Springfield community samples) on the Drive:IPIP facets. As can be expected, a smaller proportion of significant facets emerged for these criterion-based composites, since their constituent criteria were broadly but not exhaustively representative of the construct variance, compared to the criteria used in Part I. Neither self-discipline nor achievement-striving, the two facets lacking unique variance in the criterion-based composite in Part I, were among the six facets showing predictive effects in any of the three analyses conducted here. In other words, no additional predictors emerged in these analyses, which speaks to the representatives of the criterion-based composites devised and used in Part I.

Table 6.6. *Part II: Stepwise Regression Analysis Summaries for Drive:IPIP Facets Predicting the Criterion-Based Composites*

	British sample (N = 171)			Norwegian sample (N = 143)			ESCS (N = 208)		
Drive:IPIP facets	β	F	R^2_{Adj}	β	F	R^2_{Adj}	β	F	R^2_{Adj}
Model 1 (all facets)		20.31**	.60		23.04**	.65		4.78**	.18
Final model		85.97**	.60		87.23**	.65		22.18**	.17
Self-confidence				.16*					
Zest/enthusiasm/vitality	.30**			.43**					
Liveliness	.36**			.34**					
Diligence	.28**								
Ind./persev./persis.							.24**		
Joyfulness							.32**		

Note. Only beta weights for facets retained in the final models are displayed. IPIP = International Personality Item Pool; ESCS = Eugene-Springfield community sample; Ind./persev./persis. = Industriousness/perseverance/persistence.
**p* < .01. ***p* < .001.

6.5. Part III

In this part of the study, Steps 4 and 5 of FB were executed, based on the combined results of Parts I and II. A composite of Drive facets that showed predictive effects in at least one of the stepwise regressions conducted in Parts I and II was compared to the 13-facet composite in all four study samples. It was then examined whether non-predictive facets are redundant or extraneous. These analyses were conducted in all four study samples, using all seven criterion-based composites used in Parts I and II.

6.5.1. Method

Correlations of the original 13-facet composite and the modified 11-facet composite with the criterion-based composites were examined (Step 4 of FB). These amounted to a total of seven pairs of associations, which were compared using Steiger's Z test. Subsequently, the two facets identified as lacking unique common variance in Parts I and II were examined for their associations with the modified composite, in order to determine their status as redundant versus extraneous (Step 5 of FB).

6.5.2. Results and discussion

6.5.2.1. Correlations of original and modified scale composites with the criterion-based composite

Correlations of the two Drive composites with the criterion-based composite, as shown in Table 6.7, were virtually identical. The absolute difference between their associations was always less than .02, although this negligible difference reached significance in two instances—as can be expected, in favour of the modified composite. Correlations were largest for criterion-based composites that were most representative of the hypothetical facets and construct variance. These were the composite based on facet estimates and, to a lesser extent, the shared principal component of state motivation and general motivation estimates. Correlations involving the facet estimates provided by informants were lower, as would be expected, given different rating sources of facets and estimates.

Table 6.7. *Part III: Correlations of the Original and Modified Drive:IPIP Composites with the Criterion-Based Composites*

Sample (*N*)	Original scale composite	Modified scale composite	Steiger's *Z*
	Part I[a]		
Online (351)	.81	.82	-2.06*
British (233)	.73	.74	-0.43
Informants (147)	.27	.28	-1.02
Norwegian (143)	.83	.84	-0.43
	Part II[b]		
British (171)	.71	.71	0.69
Norwegian (143)	.75	.76	-2.17*
ESCS (208)	.32	.32	-0.45

Note. All coefficients are significant at $p < .001$. "Original" refers to the scale composite of all 13 facets; "Modified" refers to the 11-facet scale composite minus the self-discipline and achievement-striving facets. IPIP = International Personality Item Pool; ESCS = Eugene-Springfield community sample.
[a]Criterion-based composite derived from direct estimates of each facet on a visual analogue (Online and British samples, including informants) or Likert scale (Norwegian sample). [b]Criterion-based composite derived from state motivation composite and general motivation estimate (British and Norwegian samples) or construct-relevant criteria: depression, healthy diet, exercise, substance use/smoking, and work avoidance (ESCS).
*$p < .05$.

6.5.2.2. Correlations of self-discipline and achievement-striving with the modified scale composites and the criterion-based composite

These associations, which are displayed in Table 6.8, were within a moderate-to-strong range of r = .54 to .74, albeit somewhat lower in the Eugene-Springfield community sample. The sheer magnitude of these associations supports the notion that self-discipline and achievement-striving are redundant facets. Both share considerable variance with the construct, as operationalised by the other facets, without representation a unique part of the construct.

Table 6.8. *Part III: Correlations of Self-Discipline and Achievement-Striving with the Modified Drive:IPIP Composite*

Sample (*N*)	Self-discipline	Achievement-striving
Online (351)	.74	.71
British (233)	.70	.67
Norwegian (143)	.73	.67
ESCS (208)	.58	.54

Note. All coefficients are significant at p < .001. IPIP = International Personality Item Pool; ESCS = Eugene-Springfield community sample.

6.6. General Discussion

This investigation marks the third instalment of a series of studies examining the efficacy and utility of FB. The efficacy of FB in regards to its target focus on identifying problem facets was previously demonstrated for both redundant (**Chapter 2**) and extraneous facets (**Chapter 4**) in the context of trait EI and mindfulness, respectively. The present investigation gave further support for the method as regards identifying redundant facets, whereas no extraneous facets were apparent in the Drive:IPIP, the measure scrutinised in this research. Beyond demonstrating the efficacy of FB in the context of a new construct (general motivation), it extends confidence in its integrity considerably in two important ways, as explained next.

A key aim of this study was to examine whether the efficacy of FB is may be an artefact of common-method effects in measuring constructs and criteria. In general, variations between samples can be expected, as previously demonstrated for instances in which even the criteria are the same (**Chapter 4**) or similar (**Chapter 2**) and measured in the same way as the facets. It is, therefore, difficult to attribute any differences in predictive facets between samples completing the same criteria to the different response scales used. Here, varying the response scale format of the criteria between Likert and visual analogue scale did not seem to produce distinct results, compared to between-sample variations on Likert-scale-only data. Most of the facets with predictive effects in the two samples giving their estimates on a visual analogue scale were the same as those in the sample using a Likert measurement format for all variables. These results may alleviate any concerns regarding method effects somewhat, to the extent the Likert and visual analogue scales do not overlap entirely in terms of response bias.

As far as common-source bias is concerned, only a single facet with significant predictive effects (joyfulness) emerged from the sample involving informant-estimates as criteria. This facet was already significant in one of the other samples, in which both facets and criteria were based self-report. On the one hand, the results indicate that it was generally difficult for any facet to predict informant-based data, rendering the present results somewhat inconclusive. Yet, while the sole reliance on self-report to assess facets and criteria seems to have spawned a greater number of significant facets, integrating information ratings did not reveal any other facets with unique common variance; varying effects between different measurement methods might cast doubt on the integrity of FB, especially if there is little common ground between them.

Moreover, the informant data may offer preliminary evidence that the sole reliance on self-report is not prone to over-identifying facets as redundant or extraneous. On the contrary, using self-report data seems to give conservative estimates in regards to detecting these problem facets in the sense that it is more likely to miss redundant or extraneous facets than to classify any valid facets as problematic. In any case, it would appear to be “good” practice to integrate different measurement methods for criteria and facets, in order to minimise the risk of falsely identifying any valid facets as redundant or extraneous, and of missing any problem facets. Variations in the significance of facets seen here, although difficult to explain, support the use of multi-method approaches (Campbell & Fiske, 1959) to arrive at confident conclusions about redundant and extraneous facets.

A second major milestone achieved in the current investigation is that FB was shown to function even when all facets have a corresponding and unique criterion encompassed within the criterion-based composite. In the two preceding applications of FB, the same facets emerged as non-predictive across varying sets of both systematically selected and available criteria. Although the results were promising, it was uncertain based on these prior results whether all facets, particularly those identified as redundant or extraneous, had corresponding criteria. In the current investigation, all facets included in the measure subjected to FB were represented, at least conceptually, in the criterion-based composites used in Part I. Therefore, scale composites refined in a properly executed application of FB do not seem to achieve gains in construct validity as a result of systematic bias in the criterion-based representation of the construct variance. Rather, the presence, identification, and removal of redundant and/or extraneous facets appears to be the main reason for these improvements.

As far as the GFM is concerned, it may be reasonable to proceed in further research and applications using the Drive forms without the prima facie relevant self-discipline and achievement-striving facets. This decision would be warranted on the basis of the integrity of FB, which has been demonstrated in three studies with as many constructs. It could be further justified on the basis of the rigorous design of the current study, in which the criterion-based composites were, in theory, representative of the Drive facets and operationalised using different response scales and rating sources. Potential methodological confounds of FB were not indicated here, suggesting that the self-discipline and achievement-striving facets are redundant, relative to the other Drive facets.

6.7. Conclusions

The findings presented in the current chapter further support the efficacy and integrity of FB. The method was applied in the context of a new construct (the GFM), and identified two of its facets as redundant. A rigorous and advanced application of FB was conducted to examine alternative explanations for its effects. Specifically, a benchmark representative of all hypothetical facets was used when assessing the facets for unique construct variance. What is more, the results did not indicate any problematic common-method bias that would compromise the method's integrity. In fact, they indicated that, where facets and criteria are assessed with the same or similar methods of measurement, FB gives conservative estimates in the identification of redundant and extraneous facets.

CHAPTER 7: General Discussion

Decades ago, Cronbach and Meehl (1955) noted that there is no adequate criterion for operationally defining personality traits and other psychological constructs, which prompted their concept of construct validity and gave rise to the paradigm that governs present-day psychometrics. However, researchers continue to dwell on the level of arbitrariness involved in the process of scale construction (construct validation), often being exposed to a diversity of measurement instruments and an overall plethora of facets. FB, the psychometric method illustrated herein, is an effort towards refining multi-faceted assessment instruments in terms of construct validity, including the models on which they are based. As specified throughout this thesis, its particular aim is to identify redundant and extraneous facets. FB thereby aims to improve the level of objectivity in the construction of psychometric measures.

After summarising the findings presented in the empirical chapters of this thesis (**Chapter 2–6**), the present chapter discusses their implications in detail. It then presents a brief summary of FB, discusses considerations in its application, and explores aspects of future development. Finally, the chapter describes the limitations of the current thesis and proposes avenues for future research.

7.1. Summary of findings

Chapter 2 presented a preliminary, yet extensive application of FB in the context of trait EI, using the data from six samples. It classified four facets of a 15-facet measure of the construct as redundant. The modified scale composite not including these five facets evidenced improved construct validity in all samples. Although promising, the findings were limited in that FB was applied to available datasets; the criteria used to derive the criterion-based construct representations varied in number and types across samples. Even though the overall sample of criteria was large and diverse, it was still possible that it did not represent the construct variance fully. The two applications presented in **Chapters 4** and **6** were considerably advanced in this respect and specifically designed to examine the method.

Chapters 3 and **5** were conducted to lay the important groundwork for the applications of FB presented in **Chapters 4** and **6**, respectively. The findings of both these chapters supported the application of FB to relevant measures of the focal constructs. Specifically, **Chapter 3** showed that a core of the facets of the measures used in **Chapter 4** is related to the same dimension (conceptualised as “dispositional mindfulness”), whereas **Chapter 5** presented the development and initial validation of a measure of the GFM, which evidenced unidimensionality and, more generally, construct validity. For a more detailed discussion of these findings, the reader is invited to consult the respective chapters directly.

In **Chapter 4**, the criteria used to examine FB were selected to represent the focal construct (mindfulness) comprehensively. The most commonly used validation criteria were administered with multi-faceted mindfulness scales to several samples, and the results were cross-validated in another sample, using the least common criteria. In all instances, the same facet did not occupy unique construct variance in the criterion-based composite, and its removal led to significant improvements in construct validity. This facet, which has generated problematic results in the literature (see **Chapter 3**), was identified as extraneous to the measures concerned as well as to the mainstream conceptualisation of mindfulness. Moreover, a joint application of FB to the three independent measures scrutinised in **Chapter 3** showed that the main measure, and at least one of its facets, is not fully representative; a conceptually similar subscale of a different measure occupied unique variance in the criterion-based composite. These findings considerably supported the integrity and efficacy of FB. At the same time, it was not possible to rule out shared method effects as an alternative explanation of the thitherto promising findings.

Method effects were examined in **Chapter 6**, in which FB was applied to a relatively novel construct and a new measure. Specifically, they were addressed by (a) using different response scales for samples and criteria, and (b) using informants to provide data on the criteria in some of the study samples. Although method effects were noted, particularly in comparing results based on participant-informant versus participant-only ratings, two motivation facets did not uniquely relate to the criterion-based composite in any of the samples and analyses conducted. These two facets were classified as redundant.

While one can never be sure that a measure is fully representative of the construct, another advancement of this final investigation was that a uniquely relevant criterion was measured for each facet and included in the criterion-based composite. In particular, facet estimates were used for this purpose. This last investigation of FB, therefore, alleviates any concerns of underrepresentations of particular facets considerably. In summary, all three chapters that examined FB provided unanimous support for the method's efficacy and integrity.

7.2. Implications of Facet Benchmarking

7.2.1. Direct gains

Identification and eventual removal of redundant and extraneous facets will positively impinge on the construct validity of measures, by minimising construct-unrelated variances linked to extraneous variance and unbalanced representation attributable to redundant common variance. The improvement will be apparent in the various specific and empirically testable aspects of construct validity, which are systematically compromised by redundant and extraneous facets. For example, in detecting these problem facets, FB minimises multidimensionality and increases homogeneity within scales intended to assess a single construct, the importance of which has been discussed in detail elsewhere (Smith et al., 2009; Smith & Zapolski, 2009).

Although not primarily intended for this purpose, it seems that FB can also help reveal if individual measures and even facets are fully representative, relative to other available measures and their facets, respectively. If applied jointly to multiple measures of the construct, it can help determine whether a given measure needs to be extended, by incorporating additional facets or items. Alternatively, FB may show that some measures represent no unique construct variance, or variance not already covered by other measures, thereby casting doubt on their value.

Identification and eventual removal of redundant and extraneous facets would also entail realistic benefits for psychology's applications, particularly where psychometric assessment is involved. On a general level, it would enhance the professional and social utility of a range of standardised measures, enabling more accurate assessments of individuals and prediction of their future behaviour. Failing to represent and measure a construct adequately can have serious consequences, as psychometric assessment often forms the basis of high-stakes decisions, such as clinical diagnoses, career selection, and people matching. Another benefit of identifying, and eventually removing, redundant and extraneous facets is the reduced length of psychometric measures and shorter assessment times without trade-offs (Smith et al., 2003). Naturally, these immediate gains in construct validity can lead to other desirable consequences.

7.2.2. “Big Picture” contributions

On a larger scale, FB has the potential to help minimise the inflation of facets and proliferation of measures seen in the literature of many constructs. The presence and use of multiple measures complicates the comparison and aggregation of research findings, particularly if the models underlying these measures vary. Optimising the scale-construction process by integrating FB early on can, thus, lead to more valid conclusions about constructs, especially at the earlier stages of research. Likewise, the improvements in construct validity attributable to FB will instil greater confidence in research findings based on measures that have been scrutinised by the method; it would be possible to assess and understand the construct’s relationships with other constructs and criteria more accurately much sooner. Along these lines, FB can then also help cut research costs.

In terms of psychological applications, FB has potential in advancing overall assessment accuracy, which is pivotal to decisions made based on assessment outcomes. The improved validity of psychometric measures will result in greater accuracy in differential diagnosis and in better overall understanding of the person assessed. Additionally, the reduced length of measures scrutinised by FB minimises administration times and testing fatigue. In the case of typical-performance measures, testing fatigue can manifest itself in response inaccuracy (e.g., through a sloppy response style), whereas in maximum-performance testing situations, it can lead to test scores below the examinee’s ability. In both situations, it compromises the validity of scores. In short, FB can also minimise threats of testing effects to the validity of scores and, thereby further enhance assessment accuracy.

7.3. Facet Benchmarking: Summary, Recommendations for Application, and Projected Developments

In its current form, the method can be summarised as a five-step process: (1) select a comprehensive set of construct-relevant proximate outcomes to be administered along with a comprehensive, multi-faceted assessment instrument of the target construct to multiple samples; (2) perform Principal Component Analysis of criteria to derive an alternative representation of the construct variance (criterion-based composite) and identify unrelated individual criteria; (3) using the stepwise method, regress the criterion-based composite on hypothetical facets to identify redundant and extraneous facets; (4) compare associations of the original composite and of the composite comprising only the predictive facets with criterion-based composites; and (5) examine correlations of non-predictive facets with criterion-based composites to differentiate between redundant and extraneous facets.

If not already incorporated within this five-step process, a desirable sixth step would be to cross-validate the results in other samples, by comparing the revised and original composites in samples of criteria not included in the preceding steps. These results may or may not indicate the need for further modifications (i.e., adding facets back in). All of the statistical steps and analyses can be readily performed using conventional statistics packages (e.g., SPSS). To compare the dependent correlations computed in Step 4, syntax or more convenient external programmes can be utilised. For example, an open-source tool made available by Garbin (n.d.) can be downloaded online for the purpose of calculating Steiger's *Z*.

It is important to note that, although the five-step procedure of FB is linear, it is embedded within the construct validation paradigm, which is a non-linear and never-ending process that integrates results from different approaches. In view of its benefits, FB would be ideally integrated at the early stages of construct validation, provided that a representative sample of criteria has surfaced in the literature or can be derived at that point. For constructs that already have a well-validated operationalisation, the method can be used either to refine them or, should problem facets not emerge, to increase confidence in them and their underlying models. As indicated throughout this thesis, repeated application of FB across samples of participants and criteria will increase certainty in the identification of redundant and extraneous facets for any given construct and its operational vehicles. While these steps fully encompass the basic principles in which FB is grounded, it is important to stress that the procedure should not be conceived of as a doctrine. As any good theory or method, FB may well evolve in terms of its specific procedures over time, and application guidelines can be expected to become refined.

Future developments of FB are foreseeable with regard to two of its five steps. The first concerns the process of selecting and testing criteria for deriving alternative representations of the construct variance at Steps 1 and 2. It is anticipated that, with theoretical development and repeated application of the method, more specific examples and guidelines for criteria selection will emerge. Second, while the statistical procedures employed in FB (particularly at Step 3) can identify redundant and extraneous facets, they are of limited utility in examining the relative proportions that the remaining facets occupy within the construct variance, due to intercorrelations among predictors. However, new approaches, such as relative weight analysis, and computer software (Johnson, 2000; Nimon & Oswald, 2013; Tonidandel, Lebreton, & Johnson, 2009), may be able to estimate the relative common variances occupied by facets at Step 3. This information would provide insight into the centrality of the different valid facets and further researchers' understanding of the construct. Last, while the generic problems associated with stepwise regression algorithms are of lesser threat to FB, given its multi-sample and replication requirements, additional adjustments may be reasonable and enhance the efficacy of the method further (e.g., accounting for chance effects by using different *p*-value cut-offs or effect size estimates).

Another possible niche for development concerns the comparison of the original and modified scale composites at Step 4. In particular, different methods for statistical comparison may be preferred. Steiger's *Z* used in this thesis is a difference-based test. However, equivalence tests may have added, or even superior, value in comparing the correlations concerned. A discussion of available tests for comparing correlations was recently presented by Counsell and Cribbie (2014).

7.4. Limitations of Current Thesis and Future Directions

A common theme across the constructs to which FB has been applied so far is that they represent typical, or average, characteristics of the person (i.e., personality). Moving forward, it will be crucial to examine the method's potential for different types of constructs outside the domains of typical attributes. It would be particularly worthwhile to demonstrate that FB also has efficacy within the realm of cognitive abilities, as can be expected. Accordingly, FB should be examined in the context of maximum-performance measures and relevant criteria. Moreover, FB may have utility in the refinement of measures assessing transient mental states. State measures are operationally similar to trait measures in terms of scale format, but they differ in their focus on a limited time span of only a few days or weeks. FB is theoretically applicable to most psychological constructs, provided that their measures are homogenous enough to tap into a single dimension (Smith et al., 2009; Smith & Zapolski, 2009).

A yet unaddressed issue, albeit of peripheral importance, is the external validity (e.g., criterion, convergent, discriminant, etc.) of the measures modified through FB, compared to the original, pre-FB measures. Approaches to quantifying and comparing the external aspects of construct validity have been advanced and may prove beneficial in this respect (Westen & Rosenthal, 2003). If FB-refined measures show systematically stronger correlations with measures of related constructs and weaker correlations with conceptually unrelated constructs than the original measures, then the evidence for their gains in construct validity (in particular, criterion, convergent, discriminant validity) is more convincing. If the correlations are similar, then FB is still useful in regards to optimising the length and assessment duration of unnecessarily extensive measures. The key requirement is that their correlations with external constructs are not weaker, compared to the original scales. Re-analyses of previous datasets containing objective criteria will be invaluable in addressing this point.

Although the data used and results obtained across the investigations conducted so far give no particular reason for concern, the reliable assessment of facets needs to be given due attention in future research. Given the focus of FB, the internal reliability of facets is of utmost importance to the method's integrity, since unreliability of facets means that their validity is weak, which weakens their empirical effects, compared to those the more reliable facets (e.g., at Step 3 of FB). Specifically, the role of differences in internal reliability in FB may need to be scrutinised and taken into account systematically, possibly by means of emerging and developing statistical procedures (e.g., Tonidandel et al., 2009). On the other hand, one might argue that if a facet is not measured with adequate reliability, there is no reason to keep it in a measure, in which case FB "comes in handy" too.

The utility and impact of FB will become apparent as it continues to the scrutinised and applied to the refinement of psychological assessment instruments. Simultaneously, confidence in its integrity can be expected to increase and widen across academics and psychometricians.

7.5. Conclusions

After three investigations that have produced highly encouraging results, confidence in the efficacy of FB is very high. The method has proven useful across measures of three constructs, each scrutinised against multiple criterion-based composites, derived from varying, and sometimes systematically selected, sets of criteria. In the context of its first application on a measure of trait EI (**Chapter 2**), no definitive conclusions could be drawn from the promising findings, whereas the rigorous design of the applications presented in **Chapters 4** and **6** provides solid arguments in favour of the method. Even variations in measurement format and source did not seem to impugn the method's ability to identify redundant or extraneous facets. Importantly, FB ought be situated and applied within the context and programs of construct validation. In conjunction with existing methods for scale construction, it seems to have tremendous potential in advancing psychometric research and assessment applications.

REFERENCES

Aguado, J., Luciano, J. V., Cebolla, A., Serrano-Blanco, A., Soler, J., & García-Campayo, J. (2015). Bifactor analysis and construct validity of the five facet mindfulness questionnaire (FFMQ) in non-clinical Spanish samples. *Frontiers in Psychology*, *6*, 1–14. http://doi.org/10.3389/fpsyg.2015.00404

Alexandrov, A. (2010). Characteristics of single-item measures in likert scale format. *Electronic Journal of Business Research Methods*, *8*(1), 1–12.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: Author. http://doi.org/http://dx.doi.org/10.1176/appi.books.9780890425596.893619

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, *43*(3), 335–344. http://doi.org/10.1016/j.jrp.2008.12.013

Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment*, *11*(3), 191–206. http://doi.org/10.1177/1073191104268029

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13*(1), 27–45. http://doi.org/10.1177/1073191105283504

Barnes, S., Brown, K. W., Krusemark, E., Campbell, W. K., & Rogge, R. D. (2007). The role of mindfulness in romantic relationship satisfaction and responses to relationship stress. *Journal of Marital and Family Therapy*, *33*, 482–500. http://doi.org/10.1111/j.1752-0606.2007.00033.x

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. http://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory (2nd ed.)*. San Antonio, TX: The Psychological Corporation. http://doi.org/http://dx.doi.org/10.1007/springerreference_184604

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571. http://doi.org/http://dx.doi.org/10.1001/archpsyc.1961.01710120031004

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*, 78–117. http://doi.org/10.1177/0049124187016001004

Bergomi, C., Tschacher, W., & Kupper, Z. (2012). Measuring mindfulness: First steps towards the development of a comprehensive mindfulness scale. *Mindfulness*, *4*(1), 18–32. http://doi.org/10.1007/s12671-012-0102-9

Bergomi, C., Tschacher, W., & Kupper, Z. (2013). The assessment of mindfulness with self-report measures: Existing scales and open issues. *Mindfulness*, *4*(3), 191–202. http://doi.org/http://dx.doi.org/10.1007/s12671-012-0110-9

Bernard, L. C. (2009). Consensual and behavioral validity of a measure of adaptive individual differences dimensions in human motivation. *Motivation and Emotion*, *33*(3), 303–319. http://doi.org/10.1007/s11031-009-9131-7

Bernard, L. C., Mills, M., Swenson, L., & Walsh, R. P. (2005). An evolutionary theory of human motivation. *Genetic, Social, and General Psychology Monographs*, *131*(2), 129–184. http://doi.org/10.3200/MONO.131.2.129-184

Bernard, L. C., Mills, M., Swenson, L., & Walsh, R. P. (2008). Measuring motivation multidimensionally: Development of the Assessment of Individual Motives-

Questionnaire (AIM-Q). *Assessment*, *15*(1), 16–35. http://doi.org/10.1177/1073191107306131

Bodner, T. E., & Langer, E. J. (2001). Individual differences in mindfulness: The Mindfulness/Mindlessness Scale. *Poster Presented at the 13th Annual American Psychological Society Conference*. Toronto, Canada.

Bohlmeijer, E., ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the Five Facet Mindfulness Questionnaire in depressed adults and development of a short form. *Assessment*, *18*(3), 308–20. http://doi.org/10.1177/1073191111408231

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., … Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*(4), 676–688. http://doi.org/http://dx.doi.org/10.1016/j.beth.2011.03.007

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. http://doi.org/10.1037/0033-295X.111.4.1061

Bowden, S. C., Saklofske, D. H., & Weiss, L. G. (2011). Invariance of the measurement model underlying the Wechsler Adult Intelligence Scale-IV in the United States and Canada. *Educational and Psychological Measurement*, *71*, 186–199. http://doi.org/10.1177/0013164410387382

Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, *29*(9), 1147–1158. http://doi.org/10.1177/0146167203254596

Broadbent, D. E., Cooper, P. F., Fitzgerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *The British Journal of Clinical Psychology*, *21*(Pt 1), 1–16. http://doi.org/10.1111/j.2044-8260.1982.tb01421.x

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*(4), 822–848. http://doi.org/10.1037/0022-3514.84.4.822

Brown, K. W., Ryan, R. M., & Creswell, J. D. (2007). Addressing fundamental questions about mindfulness. *Psychological Inquiry*, *18*(4), 272–281. http://doi.org/10.1080/10478400701703344

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, *39*(3), 214–227. http://doi.org/10.1037//0003-066X.39.3.214

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452–459. http://doi.org/http://dx.doi.org/10.1037//0022-3514.63.3.452

Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review*, *90*, 105–126. http://doi.org/10.1037/0033-295X.90.2.105

Byrne, B. M. (1994). *Structural Equation Modeling with EQS and EQS-Windows: Basic Concepts, Applications, and Programming*. Thousand Oaks, CA: Sage Publications, Inc.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–7. http://doi.org/10.1207/s15327752jpa4803_13

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

http://doi.org/http://dx.doi.org/10.4135/9781412961288.n79

Cardaciotto, L., Herbert, J. D., Forman, E. M., Moitra, E., & Farrow, V. (2008). The assessment of present-moment awareness and acceptance: The Philadelphia Mindfulness Scale. *Assessment, 15*(2), 204–223. http://doi.org/10.1177/1073191107311467

Carver, C. S. (1989). How should multifaceted personality constructs be tested? Issues illustrated by self-monitoring, attributional style, and hardiness. *Journal of Personality and Social Psychology*, *56*(4), 577–85. http://doi.org/http://dx.doi.org/10.1037//0022-3514.56.4.577

Carver, C. S., Voie, L. La, Kuhl, J., & Ganellen, R. J. (1988). Cognitive concomitants of depression: A further examination of the roles of generalization, high standards, and self-criticism. *Journal of Social and Clinical Psychology*, *7*, 350–365. http://doi.org/10.1521/jscp.1988.7.4.350

Cash, M., & Whittingham, K. (2010). What facets of mindfulness contribute to psychological well-being and depressive, anxious, and stress-related symptomatology? *Mindfulness*, *1*(3), 177–182. http://doi.org/10.1007/s12671-010-0023-4

Cattell, R. B., & Kline, P. (1977). *The scientific analysis of motives and motivation*. London, UK: Academic Press.

Cebolla, A., García-Palacios, A., Soler, J., Guillen, V., Baños, R., & Botella, C. (2012). Psychometric properties of the Spanish validation of the Five Facets of Mindfulness Questionnaire (FFMQ). *The European Journal of Psychiatry, 26*(2), 118–126. http://doi.org/10.4321/S0213-61632012000200005

Chadwick, P., Hember, M., Symes, J., Peters, E., Kuipers, E., & Dagnan, D. (2008). Responding mindfully to unpleasant thoughts and images: Reliability and validity of the Southampton Mindfulness Questionnaire (SMQ). *The British Journal of Clinical Psychology*, *47*(Pt 4), 451–455. http://doi.org/10.1348/014466508X314891

Chamorro-Premuzic, T. (2007). *Personality and individual differences* (2nd ed.). Glasgow, UK: British Psychological Society and Blackwell Publishing Ltd. http://doi.org/http://dx.doi.org/10.4324/9780203785218

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219–51. http://doi.org/10.1111/j.1467-6494.2011.00739.x

Christopher, M. S., Charoensuk, S., Gilbert, B. D., Neary, T. J., & Pearce, K. L. (2009). Mindfulness in Thailand and the United States: A case of apples versus oranges? *Journal of Clinical Psychology*, *65*, 590–612. http://doi.org/10.1002/jclp.20580

Christopher, M. S., & Gilbert, B. D. (2009). Incremental validity of components of mindfulness in the prediction of satisfaction with life and depression. *Current Psychology*, *29*(1), 10–23. http://doi.org/10.1007/s12144-009-9067-9

Christopher, M. S., Neuser, N. J., Michael, P. G., & Baitmangalkar, A. (2012). Exploring the psychometric properties of the Five Facet Mindfulness Questionnaire. *Mindfulness*, *3*(2), 124–131. http://doi.org/10.1007/s12671-011-0086-x

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319. http://doi.org/10.1037/1040-3590.7.3.309

Consedine, N. S., & Butler, H. F. (2014). Mindfulness, health symptoms and healthcare utilization: Active facets and possible affective mediators. *Psychology, Health &*

Medicine, *19*(4), 392–401. http://doi.org/10.1080/13548506.2013.824596

Conte, H. R., Ratto, R., & Karasu, T. B. (1996). The Psychological Mindedness Scale: Factor structure and relationship to outcome of psychotherapy. *Journal of Psychotherapy Practice & Research*, *5*(3), 250–259.

Corr, P. J., DeYoung, C. G., & McNaughton, N. (2013). Motivation and personality: A neuropsychological perspective. *Social and Personality Psychology Compass*, *7*(3), 158–175. http://doi.org/10.1111/spc3.12016

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. http://doi.org/10.1037/0021-9010.78.1.98

Costa Jr., P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa Jr., P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, *64*(1), 21–50. http://doi.org/10.1207/s15327752jpa6401_2

Costa, P. T., & McCrae, R. R. (1998). Six approaches to the explication of facet-level traits: examples from conscientiousness. *European Journal of Personality*, *12*(2), 117–134. http://doi.org/10.1002/(SICI)1099-0984(199803/04)12:2<117::AID-PER295>3.0.CO;2-C

Counsell, A., & Cribbie, R. A. (2014). Equivalence tests for comparing correlation and regression coefficients. *British Journal of Mathematical and Statistical Psychology*, *68*(2), 292–309. http://doi.org/10.1111/bmsp.12045

Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum

Associates.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. http://doi.org/10.1037/h0040957

Curtiss, J., & Klemanski, D. H. (2014). Factor analysis of the Five Facet Mindfulness Questionnaire in a heterogeneous clinical sample. *Journal of Psychopathology and Behavioral Assessment*, *36*, 683–694. http://doi.org/10.1007/s10862-014-9429-y

Davis, K. M., Lau, M. A., & Cairns, D. R. (2009). Development and preliminary validation of a trait version of the Toronto Mindfulness Scale. *Journal of Cognitive Psychotherapy*, *23*(3), 185–197. http://doi.org/10.1891/0889-8391.23.3.185

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor model of personality: First steps towards a theory-based conceptual framework, *42*, 1285–1302. http://doi.org/10.1016/j.jrp.2008.04.002

Derksen, S., & Keselman, H. J. (2011). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*(2), 265–282. http://doi.org/10.1111/j.2044-8317.1992.tb00992.x

DeYoung, C. G. (2014). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58. http://doi.org/10.1016/j.jrp.2014.07.004

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–96. http://doi.org/10.1037/0022-3514.93.5.880

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *15*(1), 71–75. http://doi.org/http://dx.doi.org/10.1207/s15327752jpa4901_13

Dishman, R. K., Ickes, W., & Morgan, W. P. (1980). Self-motivation and adherence to habitual physical activity. *Journal of Applied Social Psychology*, *10*(2), 115–132. http://doi.org/10.1111/j.1559-1816.1980.tb00697.x

DiTommaso, E., Brannen, C., & Best, L. A. (2004). Measurement and validity characteristics of the short version of the Social and Emotional Loneliness Scale for Adults. *Educational and Psychological Measurement*, *64*(1), 99–119. http://doi.org/10.1177/0013164403258450

Dolan, L. J. (1983). Validity analyses for the school attitude measures at three grade levels. *Educational and Psychological Measurement*, *43*, 295–303. http://doi.org/10.1177/001316448304300140

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–101. http://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, *91*(2), 166–74. http://doi.org/10.1080/00223890802634290

Dulewicz, V., Higgs, M., & Slaski, M. (2003). Measuring emotional intelligence: Content, construct and criterion-related validity. *Journal of Managerial Psychology*, *18*(5), 405–420. http://doi.org/10.1108/02683940310484017

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*(1), 218–232. http://doi.org/10.1037//0022-3514.72.1.218

Emmons, R. A. (1986). Personal strivings: An approach to personality and subjective well-being. *Journal of Personality and Social Psychology*, *51*, 1058–1068. http://doi.org/10.1037/0022-3514.51.5.1058

Epstein, S. (1984). A procedural note on the measurement of broad dispositions. *Journal of Personality*, *52*(4), 318–325. http://doi.org/10.1111/j.1467-6494.1984.tb00354.x

Erisman, S. M., & Roemer, L. (2011). A preliminary investigation of the process of mindfulness. *Mindfulness*, *3*(1), 30–43. http://doi.org/10.1007/s12671-011-0078-x

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. http://doi.org/10.1037/1082-989X.4.3.272

Fehnel, S. E., Bann, C. M., Hogue, S. L., Kwong, W. J., & Mahajan, S. S. (2004). The development and psychometric evaluation of the Motivation and Energy Inventory (MEI). *Quality of Life Research*, *13*, 1321–1336. http://doi.org/10.1023/B:QURE.0000037502.64077.4d

Feldman, G., & Hayes, A. (2005). Preparing for problems: A measure of mental anticipatory processes. *Journal of Research in Personality*, *39*(5), 487–516. http://doi.org/10.1016/j.jrp.2004.05.005

Feldman, G., Hayes, A., Kumar, S., Greeson, J., & Laurenceau, J.-P. (2006). Mindfulness and emotion regulation: The development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). *Journal of Psychopathology and Behavioral Assessment*, *29*(3), 177–190. http://doi.org/10.1007/s10862-006-9035-8

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical*

Psychology, *43*(4), 522–527. http://doi.org/http://dx.doi.org/10.1037/h0076760

Fink, S., Foran, K. A., Sweeney, A. C., & O'Hea, E. L. (2009). Sexual body esteem and mindfulness in college women. *Body Image*, *6*, 326–329. http://doi.org/10.1016/j.bodyim.2009.07.003

Freudenthaler, H. H., & Neubauer, A. C. (2007). Measuring emotional management abilities: Further evidence of the importance to distinguish between typical and maximum performance. *Personality and Individual Differences*, *42*, 1561–1572. http://doi.org/10.1016/j.paid.2006.10.031

Freudenthaler, H. H., Neubauer, A. C., Gabler, P., Scherl, W. G., & Rindermann, H. (2008). Testing and validating the Trait Emotional Intelligence Questionnaire (TEIQue) in a German-speaking sample. *Personality and Individual Differences*, *45*(7), 673–678. http://doi.org/10.1016/j.paid.2008.07.014

Furnham, A. (1990). The development of single trait personality theories. *Personality and Individual Differences*, *11*(9), 923–929. http://doi.org/http://dx.doi.org/10.1016/0191-8869(90)90273-t

Garbin, C. P. (n.d.). FZT Computator. Retrieved January 13, 2013, from http://psych.unl.edu/psycrs/statpage/comp.html

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, *58*, 898–915. http://doi.org/10.1177/0013164498058006003

Gardner, K. J., & Qualter, P. (2010). Concurrent and incremental validity of three trait emotional intelligence measures. *Australian Journal of Psychology*, *62*(1), 5–13. http://doi.org/10.1080/00049530903312857

Ghorbani, N., Watson, P. J., & Weathington, B. L. (2009). Mindfulness in Iran and the

United States: Cross-cultural structural complexity and parallel relationships with psychological adjustment. *Current Psychology*, *28*, 211–224. http://doi.org/10.1007/s12144-009-9060-3

Gignac, G. E. (2009). Psychometrics and the measurement of emotional intelligence. In C. Stough, D. H. Saklofske, & J. D. A. Parker (Eds.), *Assessing emotional intelligence: Theory, research, and applications* (pp. 9–40). New York, NY: Springer Science + Business Media. http://doi.org/10.1007/978-0-387-88370-0

Gill, D. L., Dzewaltowski, D. A., & Deeter, T. E. (1988). The relationship of competitiveness and achievement orientation to participation in sport and nonsport activities. *Journal of Sport Exercise Psychology*, *10*, 139–150.

Giluk, T. L. (2009). Mindfulness, Big Five personality, and affect: A meta-analysis. *Personality and Individual Differences*, *47*(8), 805–811. http://doi.org/10.1016/j.paid.2009.06.026

Goldberg, L. R. (2008). *The Eugene-Springfield Community Sample: Information available from the research participants* (Vol. 48). Eugene, OR: Oregon Research Institute.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. http://doi.org/10.1016/j.jrp.2005.08.007

Goldberg, L. R., & Strycker, L. A. (2002). Personality traits and eating habits: The assessment of food preferences in a large community sample. *Personality and Individual Differences*, *32*, 49–65. http://doi.org/10.1016/S0191-8869(01)00005-8

Grimm, K. J., & Widaman, K. F. (2012). Construct valdity. In H. Cooper, P. M. Camic,

D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Foundations, planning, measures, and psychometrics* (Vol. 1, pp. 621–642). Washington, DC: American Psychological Association. http://doi.org/10.1037/13619-033

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. http://doi.org/10.1037/0022-3514.85.2.348

Harkness, A. R., McNulty, J. L., & Ben-Porath, Y. S. (1995). The Personality Psychopathology Five (PSY-5): Constructs and MMPI-2 scales. *Psychological Assessment*, *7*(1), 104–114.

Hart, R., Ivtzan, I., & Hart, D. (2013). Mind the gap in mindfulness research: A comparative account of the leading schools of thought. *Review of General Psychology*, *17*(4), 453–466. http://doi.org/10.1037/a0035212

Hase, H. D., & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin*, *67*(4), 231–248. http://doi.org/http://dx.doi.org/10.1037/h0024421

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *3*, 238–247. http://doi.org/http://dx.doi.org/10.1037//1040-3590.7.3.238

Heidenreich, T., Ströhle, G., & Michalak, J. (2006). Achtsamkeit: Konzeptuelle Aspekte und Ergebnisse zum Freiburger Achtsamkeitsfragebogen. *Verhaltenstherapie*, *16*, 33–40. http://doi.org/10.1159/000091521

Höfling, V., Ströhle, G., Michalak, J., & Heidenreich, T. (2011). A short version of the

Kentucky Inventory of Mindfulness Skills. *Journal of Clinical Psychology*, *67*(6), 639–45. http://doi.org/10.1002/jclp.20778

Hogan, J., & Hogan, R. (1996). *Motives, Values, Preferences Inventory manual*. Tulsa, OK: Hogan Assessment Systems.

Hogan, R., & Hogan, J. (1992). *Hogan Personality Inventory manual* (2nd ed.). Tulsa, OK: Hogan Assessment Systems.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hull, J. G., Lehn, D. A., & Tedlie, J. C. (1991). A general approach to testing multifaceted personality constructs. *Journal of Personality and Social Psychology*, *61*(6), 932–45. http://doi.org/http://dx.doi.org/10.1037//0022-3514.61.6.932

Hurt, R. D., Morse, R. M., & Swenson, W. M. (1980). Diagnosis of alcoholism with a self-administered alcoholism screening test: Results with 1,002 consecutive patients receiving general examinations. *Mayo Clinic Proceedings*, *55*(6), 365–70.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90–98. http://doi.org/10.1016/j.jcps.2009.09.003

Jackson, D. N. (1984). *Personality Research Form manual*. Port Huron, MI: Research Psychologists Press.

Jamieson, G. A. (2005). The Modified Tellegen Absorption Scale: A clearer window on the structure and meaning of absorption. *Australian Journal of Clinical and Experimental Hypnosis*, *33*(2), 119–139.

John, O. P., & Benet-Martinez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339–369). New York, NY: Cambridge University Press. http://doi.org/http://dx.doi.org/10.1017/cbo9780511996481

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461–494). New York, NY: Guilford Press. http://doi.org/http://dx.doi.org/10.5860/choice.45-2315

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York, NY: Guilford Press.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*(1), 37–41. http://doi.org/10.1207/S15327906MBR3501

Jordan, P., Lawrence, S., Emotional, W., & Profile, I. (2009). Emotional intelligence in teams: Development and initial validation of the short version of the Workgroup Emotional Intelligence Profile (WEIP-S). *Journal of Management*, *15*(4), 452–469. http://doi.org/http://dx.doi.org/10.5172/jmo.15.4.452

Kabat-Zinn, J. (1994). *Wherever you go, there you are: Mindfulness meditation in everyday life*. New York, NY: Hyperion.

Kanfer, R., & Ackerman, P. L. (2000). Individual differences in work motivation: Further explorations of a trait framework. *Applied Psychology*, *49*(3), 470–482.

http://doi.org/10.1111/1464-0597.00026

Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., & Steger, M. F. (2009). The Curiosity and Exploration Inventory-II: Development, factor structure, and psychometrics. *Journal of Research in Personality*, *43*(6), 987–998. http://doi.org/10.1016/j.jrp.2009.04.011

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, *8*(3), 305–321. http://doi.org/10.1037/1082-989X.8.3.305

Kohls, N., Sauer, S., & Walach, H. (2009). Facets of mindfulness – Results of an online study investigating the Freiburg Mindfulness Inventory. *Personality and Individual Differences*, *46*(2), 224–230. http://doi.org/10.1016/j.paid.2008.10.009

Langer, E. J. (1989). *Mindfulness*. Cambridge, MA: Da Capo Press.

Lau, M. A., Bishop, S. R. S., Buis, T., Anderson, N. D., Carlson, L., Carmody, J., & Segal, Z. (2006). The Toronto Mindfulness Scale: Development and validation. *Journal of Clinical Psychology*, *62*(12), 1445–1467. http://doi.org/10.1002/jclp

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the Self-Monitoring Scale. *Journal of Personality and Social Psychology*, *46*, 1349–1364. http://doi.org/http://dx.doi.org/10.1037//0022-3514.46.6.1349

Liem, G. A. D., & Martin, A. J. (2012). The Motivation and Engagement Scale: Theoretical framework, psychometric properties, and applied yields. *Australian Psychologist*, *47*, 3–13. http://doi.org/10.1111/j.1742-9544.2011.00049.x

Little, B. R. (1983). Personal projects: A rationale and method for investigation. *Environment and Behavior*, *15*, 273–309. http://doi.org/10.1177/0013916583153002

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment, and personality: A study of 850 sets of twins*. Austin, TX: University of Texas Press.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694. http://doi.org/10.2466/pr0.1957.3.3.635

Loo, R. (2002). A caveat on using single-item versus multiple-item scales. *Journal of Managerial Psychology*, *17*(1), 68–75. http://doi.org/10.1108/02683940210415933

López-Ibor Aliño, J. J., Pérez Urdaníz, A., & Rubio Larrosa, V. (1996). *Examen internacional de los trastornos de la personalidad, módulo D.S.M.-IV [International Personality Disorder Examination (IPDE) - DSM-IV module]*. Madrid: Meditor.

Loranger, A. W., Janca, A., & Sartorius, N. (1997). *Assessment and diagnosis of personality disorders: The ICD-10 international personality disorder examination (IPDE)*. Cambridge, UK: Cambridge University Press.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*(2), 137–155. http://doi.org/10.1023/A:1006824100041

Martins, A., Ramalho, N., & Morin, E. (2010). A comprehensive meta-analysis of the relationship between emotional intelligence and health. *Personality and Individual Differences*, *49*(6), 554–564. http://doi.org/10.1016/j.paid.2010.05.029

Martskvishvili, K., Arutinov, L., & Mestvirishvili, M. (2011). *A psychometric investigation of the Georgian version of the Trait Emotional Intelligence Questionnaire*. Poster presented at the meeting of the International Society for the Study of Individual Differences, London, UK.

Martskvishvili, K., Arutinov, L., & Mestvirishvili, M. (2013). A psychometric investigation of the Georgian version of the Trait Emotional Intelligence Questionnaire. *European Journal of Psychological Assessment*, *29*(2), 84–88. http://doi.org/10.1027/1015-5759/a000135

Massey, E. K., Gebhardt, W. A., & Garnefski, N. (2008). Adolescent goal content and pursuit: A review of the literature from the past 16 years. *Developmental Review*, *28*(4), 421–460. http://doi.org/10.1016/j.dr.2008.03.002

Mayer, J. D., Faber, M. A., & Xu, X. (2007). Seventy-five years of motivation measures (1930–2005): A descriptive analysis. *Motivation and Emotion*, *31*(2), 83–103. http://doi.org/10.1007/s11031-007-9060-2

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence* (Vol. 23, pp. 3–31). New York, NY: Basic Books. http://doi.org/10.1177/1066480710387486

Mayer, J. L., & Farmer, R. F. (2010). The development and psychometric evaluation of a new measure of dissociative activities. *Journal of Personality Assessment*, *80*(2), 185–196. http://doi.org/http://dx.doi.org/10.1207/s15327752jpa8002_07

McClelland, D. C. (1965). Toward a theory of motive acquisition. *The American Psychologist*, *20*, 321–333. http://doi.org/10.1037/h0022225

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of cognitive abilities: Past, present, and future. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–182). New York, NY: Guilford Press. http://doi.org/http://dx.doi.org/10.1080/02783190802201986

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning.

American Psychologist, *50*, 741–749. http://doi.org/10.1037/0003-066X.50.9.741

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *28*(6), 487–495. http://doi.org/http://dx.doi.org/10.1016/0005-7967(90)90135-6

Mikolajczak, M., Luminet, O., Leroy, C., & Roy, E. (2007). Psychometric properties of the Trait Emotional Intelligence Questionnaire: Factor structure, reliability, construct, and incremental validity in a French-speaking population. *Journal of Personality Assessment*, *88*(3), 338–353. http://doi.org/http://dx.doi.org/10.1080/00223890701333431

Mikolajczak, M., Luminet, O., & Menil, C. (2006). Predicting resistance to stress: Incremental validity of trait emotional intelligence over alexithymia and optimism. *Psicothema*, *18 Suppl*, 79–88.

Mikolajczak, M., Roy, E., Verstrynge, V., & Luminet, O. (2009). An exploration of the moderating effect of trait emotional intelligence on memory and attention in neutral and stressful conditions. *British Journal of Psychology*, *100*(Pt 4), 699–715. http://doi.org/10.1348/000712608X395522

Motivation. (n.d.-a). In *Oxford Dictionaries*.

Motivation. (n.d.-b). In *Business Dictionary*.

Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.

Mussel, P. (2013). Intellect: A theoretical framework for personality traits related to intellectual achievements. *Journal of Personality and Social Psychology*, *104*(5), 885–906. http://doi.org/http://dx.doi.org/10.1037/a0031918

Neale-Lorello, D., & Haaga, D. A. F. (2015). The “observing” facet of mindfulness moderates stress/symptom relations only among meditators. *Mindfulness*, Advance online publication. http://doi.org/10.1007/s12671-015-0396-5

Nimon, K. F., & Oswald, F. L. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods*, *16*(4), 650–674. http://doi.org/10.1177/1094428113493929

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw Hill.

Nygård, R., & Gjesme, T. (1973). Assessment of achievement motives: Comments and suggestions. *Scandinavian Journal of Educational Research*, *17*, 39–46. http://doi.org/10.1080/0031383730170104

Ong, J. C., Shapiro, S. L., & Manber, R. (2009). Mindfulness meditation and cognitive behavioral therapy for insomnia: A naturalistic 12-month follow-up. *Explore: The Journal of Science and Healing*, *5*, 30–36. http://doi.org/10.1016/j.explore.2008.10.004

Osuch, E. A., Noll, J. G., & Putnam, F. W. (1999). The motivations for self-injury in psychiatric inpatients. *Psychiatry*, *62*, 334–346.

Park, T., Reilly-Spong, M., & Gross, C. R. (2013). Mindfulness: A systematic review of instruments to measure an emergent patient-reported outcome (PRO). *Quality of Life Research*, *22*(10), 2639–2659. http://doi.org/10.1007/s11136-013-0395-8

Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality*, *66*, 525–554. http://doi.org/10.1111/1467-6494.00023

Pedhazur, E. J. (1997). *Multiple-regression in behavioral research: Explanation and prediction* (3rd ed.). New York, NY: Harcourt Brace.

http://doi.org/10.2307/2285468

Pelletier, L. G., Pelletier, L. G., Tuson, K. M., Tuson, K. M., Green-Demers, I., Green-Demers, I., … Beaton, A. M. (1998). Why are you doing things for the environment? The Motivation Toward the Environment Scale (MTES). *Journal of Applied Social Psychology*, *28*, 437–468. http://doi.org/10.1111/j.1559-1816.1998.tb01714.x

Petrides, K. V. (2009). *Technical manual for the Trait Emotional Intelligence Questionnaires (TEIQue; 1st edition, 4th printing)*. London, UK: London Psychometric Laboratory.

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality*, *15*(6), 425–448. http://doi.org/10.1002/per.416

Petrides, K. V., Mason, M., & Sevdalis, N. (2011). Preliminary validation of the construct of trait social intelligence. *Personality and Individual Differences*, *50*(6), 874–877. http://doi.org/10.1016/j.paid.2010.12.029

Petrides, K. V., Pérez-González, J. C., & Furnham, A. (2007). On the criterion and incremental validity of trait emotional intelligence. *Cognition & Emotion*, *21*(1), 26–55. http://doi.org/10.1080/02699930601038912

Petrides, K. V., Pita, R., & Kokkinaki, F. (2007). The location of trait emotional intelligence in personality factor space. *British Journal of Psychology*, *98*(Pt 2), 273–89. http://doi.org/10.1348/000712606X120618

Pirson, M., Langer, E. J., Bodner, T., & Zilcha, S. (2012). The development and validation of the Langer Mindfulness Scale - enabling a socio-cognitive perspective of mindfulness in organizational contexts. *SSRN Electronic Journal*, 1–54. http://doi.org/10.2139/ssrn.2158921

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. http://doi.org/10.1037/0021-9010.88.5.879

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. http://doi.org/http://dx.doi.org/10.1177/014662167700100306

Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the Self-Compassion Scale. *Clinical Psychology & Psychotherapy*, *18*(3), 250–255. http://doi.org/10.1002/cpp.702

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor & Francis Group. http://doi.org/http://dx.doi.org/10.4324/9780203841624

Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology*, *8*, 179–193. http://doi.org/10.1037/1089-2680.8.3.179

Reiss, S., & Havercamp, S. M. (1998). Toward a comprehensive assessment of fundamental motivation: Factor structure of the Reiss Profiles. *Psychological Assessment*, *10*(2), 97–106. http://doi.org/10.1037//1040-3590.10.2.97

Revelle, W., Wilt, J., & Condon, D. M. (2011). Individual differences and differential psychology: A brief history and prospect. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 1–38). Oxford, UK: Blackwell Publishing Ltd. http://doi.org/10.1002/9781444343120.ch1

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the GLB:

Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154. http://doi.org/10.1007/s11336-008-9102-z

Roger, D., Jarvis, G., & Najarian, B. (1993). Detachment and coping: The construction and validation of a new scale for measuring coping strategies. *Personality and Individual Differences*, *15*(6), 619–626. http://doi.org/10.1016/0191-8869(93)90003-L

Roger, D., & Najarian, B. (1989). The construction and validation of a new scale for measuring emotion control. *Personality and Individual Differences*, *10*(8), 845–853. http://doi.org/http://dx.doi.org/10.1016/0191-8869(89)90020-2

Saklofske, D. H., Hildebrand, D. K., & Gorsuch, R. L. (2000). Replication of the factor structure of the Wechsler Adult Intelligence Scale--Third Edition with a Canadian sample. *Psychological Assessment*, *12*, 436–439. http://doi.org/10.1037/1040-3590.12.4.436

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, and health* (pp. 125–154). Washington, DC: American Psychological Association. http://doi.org/10.1016/0147-1767(78)90036-6

Sandín, B., Chorot, P., Lostao, L., Joiner, T. E., Santed, M. A., & Valiente, R. M. (1999). Escalas panas de afecto positivo y negativo: Validación factorial y convergencia transcultural [The PANAS scales of positive and negative affect: Factor analytic validation and cross-cultural convergence]. *Psicothema*, *11*(1), 37–51.

Sanz, J., Perdigón, A. L., & Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-ll (BDI-II): 2. Propiedades psicométricas en población general [The spanish adaptation of Beck's Depression Inventory-II (BDI-II): 2. Psychometric properties in the general population]. *Clínica Y Salud*, *14*(3), 249–280.

Sauer, S., Walach, H., Schmidt, S., Hinterberger, T., Lynch, S., Büssing, A., & Kohls, N. (2012). Assessment of mindfulness: Review on state of the art. *Mindfulness*, *4*(1), 3–17. http://doi.org/10.1007/s12671-012-0122-5

Schmidt, F. L., & Hunter, J. E. (1992). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science*, *1*(3), 89–92. http://doi.org/10.1111/1467-8721.ep10768758

Schönbrodt, F. D., & Gerstenberg, F. X. R. (2012). An IRT analysis of motive questionnaires: The Unified Motive Scales. *Journal of Research in Personality*, *46*(6), 725–742. http://doi.org/10.1016/j.jrp.2012.08.010

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, *25*(2), 167–177. http://doi.org/10.1016/S0191-8869(98)00001-4

Segal, Z., Teasdale, J., & Williams, M. (2002). *Mindfulness-based cognitive therapy for depression*. New York, NY: Guilford Press.

Seifriz, J. J., Duda, J. L., & Chi, L. K. (1992). The relationship of perceived Motivational climate to instrinsic motivation and beliefs about success in basketball. *Journal of Sport & Exercise Psychology*, *14*, 375–391.

Shah, J. Y., & Gardner, W. L. (2008). *Handbook of motivation science*. New York, NY: Guilford Press.

Shill, M. A., & Lumley, M. A. (2002). The Psychological Mindedness Scale: Factor structure, convergent validity and gender in a non-psychiatric sample.

Psychology and Psychotherapy, *75*(Pt 2), 131–50. http://doi.org/http://dx.doi.org/10.1348/147608302169607

Siegling, A. B., Saklofske, D. H., Vesely, A. K., & Nordstokke, D. W. (2012). Relations of emotional intelligence with gender-linked personality: Implications for a refinement of EI constructs. *Personality and Individual Differences*, *52*(7), 776–781. http://doi.org/10.1016/j.paid.2012.01.003

Siegling, A. B., Vesely, A. K., Petrides, K. V., & Saklofske, D. H. (2015). Incremental validity of the Trait Emotional Intelligence Questionnaire–Short Form (TEIQue–SF). *Journal of Personality Assessment*, Advance online publication. http://doi.org/10.1080/00223891.2015.1013219

Siegling, A. B., Vesely, A. K., & Saklofske, D. H. (2013). Advancing the trait EI content domain: Further evidence for the distinctiveness of interpersonal facets. *Personality and Individual Differences*, *54*(1), 81–86. http://doi.org/10.1016/j.paid.2012.08.010

Simms, L. J., & Watson, D. (2007). The construct validation approach to personality scale construction. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality* (pp. 240–258). New York, NY: Guilford. http://doi.org/http://dx.doi.org/10.5860/choice.45-2315

Smith, G. T. (2005a). On construct validity: Issues of method and measurement. *Psychological Assessment*, *17*(4), 396–408. http://doi.org/10.1037/1040-3590.17.4.396

Smith, G. T. (2005b). On the complexity of quantifying construct validity. *Psychological Assessment*, *17*, 413–414. http://doi.org/10.1037/1040-3590.17.4.413

Smith, G. T., Fischer, S., & Fister, S. M. (2003). Incremental validity principles in test construction. *Psychological Assessment*, *15*(4), 467–77. http://doi.org/10.1037/1040-3590.15.4.467

Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 300–308. http://doi.org/10.1037/1040-3590.7.3.300

Smith, G. T., McCarthy, D. M., & Zapolski, T. C. B. (2009). On the value of homogeneous constructs for construct validation, theory testing, and the description of psychopathology. *Psychological Assessment*, *21*(3), 272–284. http://doi.org/10.1037/a0016699

Smith, G. T., & Zapolski, T. C. B. (2009). Construct validation of personality measures. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 81–88). Oxford, UK: Oxford University Press. http://doi.org/http://dx.doi.org/10.1093/oxfordhb/9780195366877.001.0001

Spielberger, C. D., Gorsuch, R. L., Lushene, P. R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Stice, E., Telch, C. F., & Rizvi, S. L. (2000). Development and validation of the Eating Disorder Diagnostic Scale: A brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychological Assessment*, *12*(2), 123–131. http://doi.org/10.1037/1040-3590.12.2.123

Stone, L. L., Otten, R., Ringlever, L., Hiemstra, M., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2013). The parent version of the strengths and difficulties questionnaire: Omega as an alternative to alpha and a test for measurement invariance. *European Journal of Psychological Assessment*, *29*, 44–50. http://doi.org/10.1027/1015-5759/a000119

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*, 1–25. http://doi.org/10.1146/annurev.clinpsy.032408.153639

Swami, V., Begum, S., & Petrides, K. V. (2010). Associations between trait emotional intelligence, actual–ideal weight discrepancy, and positive body image. *Personality and Individual Differences*, *49*(5), 485–489. http://doi.org/10.1016/j.paid.2010.05.009

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson. http://doi.org/10.1037/022267

Tapia, M., & Marsh, G. E. (2006). A validation of the emotional intelligence inventory. *Psicothema*, *18 Suppl*, 55–8.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *The Journal of Applied Psychology*, *88*(3), 500–517. http://doi.org/10.1037/0021-9010.88.3.500

Tett, R. P., Fox, K. E., & Wang, A. (2005). Development and validation of a self-report measure of emotional intelligence as a multidimensional trait domain. *Personality and Social Psychology Bulletin*, *31*(7), 859–88. http://doi.org/10.1177/0146167204272860

Tonidandel, S., Lebreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods*, *14*(4), 387–99. http://doi.org/10.1037/a0017735

Tran, U. S., Cebolla, A., Glück, T. M., Soler, J., Garcia-Campayo, J., & von Moy, T. (2014). The serenity of the meditating mind: A cross-cultural psychometric study on a two-factor higher order structure of mindfulness, its effects, and mechanisms related to mental health among experienced meditators. *PLoS ONE*, *9*(10), e110192. http://doi.org/10.1371/journal.pone.0110192

Tran, U. S., Glück, T. M., & Nader, I. W. (2013). Investigating the Five Facet Mindfulness Questionnaire (FFMQ): Construction of a short form and evidence of a two-factor higher order structure of mindfulness. *Journal of Clinical Psychology*, *69*(9), 951–965. http://doi.org/10.1002/jclp.21996

Trapnell, P. D., & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology*, *76*(2), 284–304. http://doi.org/http://dx.doi.org/10.1037//0022-3514.76.2.284

Uva, M. C. D. S., Timary, P. De, Cortesi, M., Mikolajczak, M., Blicquy, P. D. R. De, & Luminet, O. (2010). Moderating effect of emotional intelligence on the role of negative affect in the motivation to drink in alcohol-dependent subjects undergoing protracted withdrawal. *Personality and Individual Differences, 48*(1), 16–21. http://doi.org/10.1016/j.paid.2009.08.004

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, *52*(4), 1003–1017. http://doi.org/10.1177/0013164492052004025

Van der Linden, D., Tsaousis, I., & Petrides, K. V. (2012). Overlap between General Factors of Personality in the Big Five, Giant Three, and trait emotional intelligence. *Personality and Individual Differences*, *53*(3), 175–179. http://doi.org/10.1016/j.paid.2012.03.001

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*(2), 231–251. http://doi.org/10.1037/1082-989X.3.2.231

Vidrine, J. I., Businelle, M. S., Cinciripini, P., Li, Y., Marcus, M. T., Waters, A. J., … Wetter, D. W. (2009). Associations of mindfulness with nicotine dependence, withdrawal, and agency. *Substance Abuse*, *30*(4), 318–27. http://doi.org/10.1080/08897070903252973

Vujanovic, A. A., Bonn-Miller, M. O., Bernstein, A., McKee, L. G., & Zvolensky, M. J. (2010). Incremental validity of mindfulness skills in relation to emotional dysregulation among a young adult community sample. *Cognitive Behaviour Therapy*, *39*(3), 203–13. http://doi.org/10.1080/16506070903441630

Walach, H., Buchheld, N., Buttenmüller, V., Kleinknecht, N., & Schmidt, S. (2006). Measuring mindfulness—the Freiburg Mindfulness Inventory (FMI). *Personality and Individual Differences*, *40*(8), 1543–1555. http://doi.org/10.1016/j.paid.2005.11.025

Watson, D. (2012). Objective tests as instruments of psychological theory and research. In H. Cooper (Ed.), *Handbook of research methods in psychology: Foundations, planning, measures, and psychometrics* (Vol. 1, pp. 349–369). New York, NY: American Psychological Association. http://doi.org/10.1037/13619-019

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect – The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. http://doi.org/10.1037/0022-3514.54.6.1063

Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality*, *62*(4), 616–640. http://doi.org/http://dx.doi.org/10.1111/j.1467-6494.1994.tb00311.x

Weiss, D. J., Dawis, R. V, England, G. W., & Lofquist, L. H. (1967). *Manual for the Minnesota Satisfaction Questionnaire*. Minneapolis, MIN: Industrial Relations

Center, University of Minnesota.

Weissman, A. N., & Beck, A. T. (1978). *Development and validation of the Dysfunctional Attitudes Scale: A preliminary investigation.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, *84*(3), 608–618. http://doi.org/10.1037/0022-3514.84.3.608

Williams, M. J., Dalgleish, T., Karl, A., & Kuyken, W. (2014). Examining the factor structures of the Five Facet Mindfulness Questionnaire and the Self-Compassion Scale. *Psychological Assessment*, *26*(2), 407–18. http://doi.org/10.1037/a0035566

Ziegler, M. (2014a). Comments on item selection procedures. *European Journal of Psychological Assessment*, *30*(1), 1–2. http://doi.org/10.1027/1015-5759/a000196

Ziegler, M. (2014b). Stop and state your intentions! Let's not forget the ABC of test construction. *European Journal of Psychological Assessment*, *30*(4), 239–242. http://doi.org/10.1027/1015-5759/a000228

Ziegler, M., Booth, T., & Bensch, D. (2013). Getting entangled in the nomological net: Thoughts on validity and conceptual overlap. *European Journal of Psychological Assessment*, *29*(3), 157–161. http://doi.org/10.1027/1015-5759/a000173

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, *69*(4), 548–565. http://doi.org/10.1177/0013164408324469

Ziegler, M., Danay, E., Schoelmreich, F., & Buehner, M. (2010). Predicting academic success with the big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality*, *24*(4), 341–355. http://doi.org/10.1002/per

Ziegler, M., Kemper, C. J., & Lenzner, T. (2015). The issue of fuzzy concepts in test construction and possible remedies. *European Journal of Psychological Assessment*, *31*(1), 1–4. http://doi.org/10.1027/1015-5759/a000255

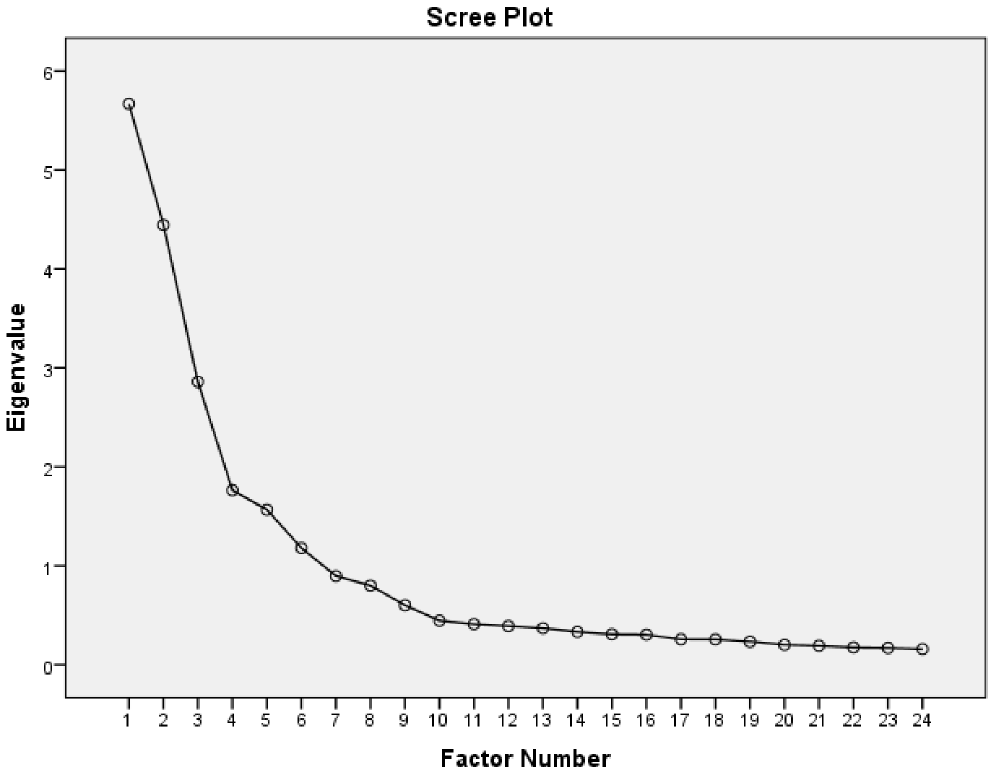
Ziegler, M., Schmukle, S., Egloff, B., & Bühner, M. (2010). Investigating measures of achievement motivation(s). *Journal of Individual Differences*, *31*(1), 15–21. http://doi.org/10.1027/1614-0001/a000002

Zuckerman, M., & Allison, S. N. (1976). An objective measure of fear of success: construction and validation. *Journal of Personality Assessment*, *40*, 422–430. http://doi.org/10.1207/s15327752jpa4004_12a

Zuckerman, M., Joireman, J., Kraft, M., & Kuhlman, D. M. (1999). Where do motivational and emotional traits fit within three factor models of personality? *Personality and Individual Differences*, *26*, 487–504. http://doi.org/10.1016/S0191-8869(98)00161-5

APPENDICES

Appendix 1: Scree plot and parallel analysis described in Chapter 3.3.2.3



Run MATRIX procedure:

PARALLEL ANALYSIS:

PAF/Common Factor Analysis & Random Normal Data Generation

Specifications for this Run:
Ncases 172
Nvars 24
Ndatsets 5000
Percent 95

Raw Data Eigenvalues, & Mean & Percentile Random Data Eigenvalues

Root	Raw Data	Means	Prcntyle
1.000000	7.144713	.904530	1.044203
2.000000	3.594540	.770957	.874777
3.000000	2.635003	.671193	.758880
4.000000	1.349201	.586330	.664487
5.000000	1.201928	.510452	.580658
6.000000	.752429	.442406	.506161
7.000000	.506539	.378856	.441415
8.000000	.438883	.319025	.376766
9.000000	.117404	.262787	.318708
10.000000	.026912	.208256	.260764
11.000000	.014079	.156342	.207154
12.000000	.012014	.107185	.153928
13.000000	.002064	.059425	.103392
14.000000	-.022413	.013663	.056182
15.000000	-.030906	-.030889	.009425
16.000000	-.046288	-.073375	-.035085
17.000000	-.057911	-.115374	-.080821
18.000000	-.087056	-.155713	-.121180
19.000000	-.099680	-.196020	-.163551
20.000000	-.118868	-.235908	-.204971
21.000000	-.125898	-.275020	-.244101
22.000000	-.143869	-.315076	-.285583
23.000000	-.164668	-.357937	-.326222
24.000000	-.187089	-.407383	-.372461

Appendix 2: Instructions given to raters of the International Personality Item Pool facets (see Chapter 5.3.2.1)

Below is a set of characteristics that were identified as potential general indicators of motivation (i.e., regardless of a person's circumstances, states, goals, desires, interests, motives, etc.)

If you don't agree that any of these characteristics is indicative of motivation or are not sure, please indicate this using the scale provided (if you agree, no response is needed since 'yes' is pre-selected).

Motivation has been defined in various ways. One definition I found is "Desire and energy in people to be continually interested and committed to a job, role or subject, or to make an effort to attain a goal."   Another is "Desire or willingness to do something."

For the characteristics below, please think of motivation in a generic sense (i.e., <u>not</u> tied to any particular circumstances, states, goals, desires, interests, motives, etc.). The question is whether the characteristics are general indicators of motivation. If you are unsure about the meaning of a characteristic, please look it up quickly (e.g., at www.dictionary.com).

Appendix 3: International Personality Item Pool facets and items used to represent and measure the General Factor of Motivation (see Chapter 5.4.1.2)

Self-confidence

+ keyed	Think highly of myself.
	Know immediately what to do.

– keyed	Have a low opinion of myself.
	Am easily intimidated.
	Feel threatened easily.

Temperance

+ keyed	Rarely overindulge.
	Keep my promises.

– keyed	Change my mood a lot.
	Am guided by my moods.
	Am preoccupied with myself.
	Grumble about things.
	Suddenly lose interest.
	Say inappropriate things.
	Love to come up with objections.

Zest/enthusiasm/vitality

+ keyed	Prefer to participate fully rather than view life from the sidelines.
	Don't approach things halfheartedly.
	Love what I do.
	Look forward to each new day.
	Can’t wait to get started on a project.
	Can hardly wait to see what life has in store for me in the years ahead.
	Awaken with a sense of excitement about the day's possibilities.

- keyed	Dread getting up in the morning.
	Don't have much energy.

Valor/bravery/courage

+ keyed	Have taken frequent stands in the face of strong opposition.
	Don't hesitate to express an unpopular opinion.
	Call for action while others talk.
	Can face my fears.
	Speak up in protest when I hear someone say mean things.
	Am a brave person.

Liveliness

+keyed Maintain high energy throughout the day.
Have great stamina.
Am usually active and full of energy.
Smile a lot.
Feel healthy and vibrant most of the time.
Laugh a lot.
Feel that I have a lot of inner strength.

- keyed Tire out quickly.

Insight

+ keyed Come up with something new.
Throw a new light on the situation.
Come up with alternatives.
Put a new perspective on things.
Have a vivid imagination.

– keyed Consider myself an average person.

Initiative

+ keyed Get things done quickly.
Get to work at once.
Finish tasks quickly.
Start tasks right away.

– keyed Put off unpleasant tasks.

Diligence

+keyed Push myself very hard to succeed.
Get started quickly on doing a job.
Am exacting in my work.

- keyed Stop when work becomes too difficult.
Quickly lose interest in the tasks I start.

Deliberateness

+ keyed Choose my words with care.
Take care of my own affairs.
Remain calm under pressure.

– keyed Like to act on a whim.
Rush into things.
Act quickly without thinking.

Competitive

+ keyed	Accept challenging tasks.
	Am good at many things.
– keyed	Am not highly motivated to succeed.
	Do just enough work to get by.
	Undertake few things on my own.

Experience-seeking

+ keyed	Prefer variety to routine.
	Seek adventure.
	Try out new things.
– keyed	Would describe my experiences as somewhat dull.
	Dislike new foods.

Generates ideas

+ keyed	Am full of ideas.
	Have excellent ideas.
	Quickly think up new ideas.
– keyed	Do not have a good imagination.
	Can't come up with new ideas.

Prudence

+keyed	Avoid mistakes.
	Make plans and stick to them.
	Do things according to a plan.
- keyed	Do things without thinking of the consequences.
	Act impulsively when something is bothering me.
	Make careless mistakes.
	Make a fool of myself.

Resourcefulness

+ keyed	Can handle complex problems.
	Face problems directly.
	Formulate ideas clearly.
– keyed	Wait for others to lead the way.
	Can't make up my mind.
	Panic easily.

Self-control

+ keyed   Am not easily affected by my emotions.
Never spend more than I can afford.
Experience very few emotional highs and lows.

– keyed   Act wild and crazy.
Demand attention.
Do crazy things.
Use flattery to get ahead.
Use swear words.
Make a lot of noise.

Ind./persev./persis.

+ keyed   Don't quit a task before it is finished.
Am a goal-oriented person.
Finish things despite obstacles in the way.
Am a hard worker.
Don't get sidetracked when I work.

- keyed   Don't finish what I start.
Do not tend to stick with what I decide to do.

Activity-level

+ keyed   Can manage many things at the same time.
Am always busy.
Do a lot in my spare time.
Am always on the go.
Accomplish a lot of work.

– keyed   Have a slow pace to my life.

Adventurousness

+ keyed   Like to visit new places.
Interested in many things.
Like to begin new things.

– keyed   Prefer to stick with things that I know.
Dislike changes.
Am a creature of habit.
Am attached to conventional ways.

Self-discipline

+ keyed Get chores done right away.
Am always prepared.
Carry out my plans.

– keyed Waste my time.
Postpone decisions.

Achievement-striving

+ keyed Go straight for the goal.
Turn plans into actions.
Do more than what's expected of me.
Set high standards for myself and others.
Demand quality.

– keyed Put little time and effort into my work.

Hope/optimism

+keyed Look on the bright side.
Can find the positive in what seems negative to others.
Remain hopeful despite challenges.
Will succeed with the goals I set for myself.
Think about what is good in my life when I feel down.

-keyed Expect the worst.
Have no plan for my life five years from now.
Am not confident that my way of doing things will work out for the best.

Competence

+ keyed Like to solve complex problems.
Can perform a wide variety of tasks.
Know how to apply my knowledge.
Meet challenges.

– keyed Don't put my mind on the task at hand.
Don't see things through.

Self-efficacy

+ keyed	Excel in what I do.
Handle tasks smoothly.
Come up with good solutions.
Know how to get things done.

– keyed	Misjudge situations.
Don't understand things.
Don't see the consequences of things.

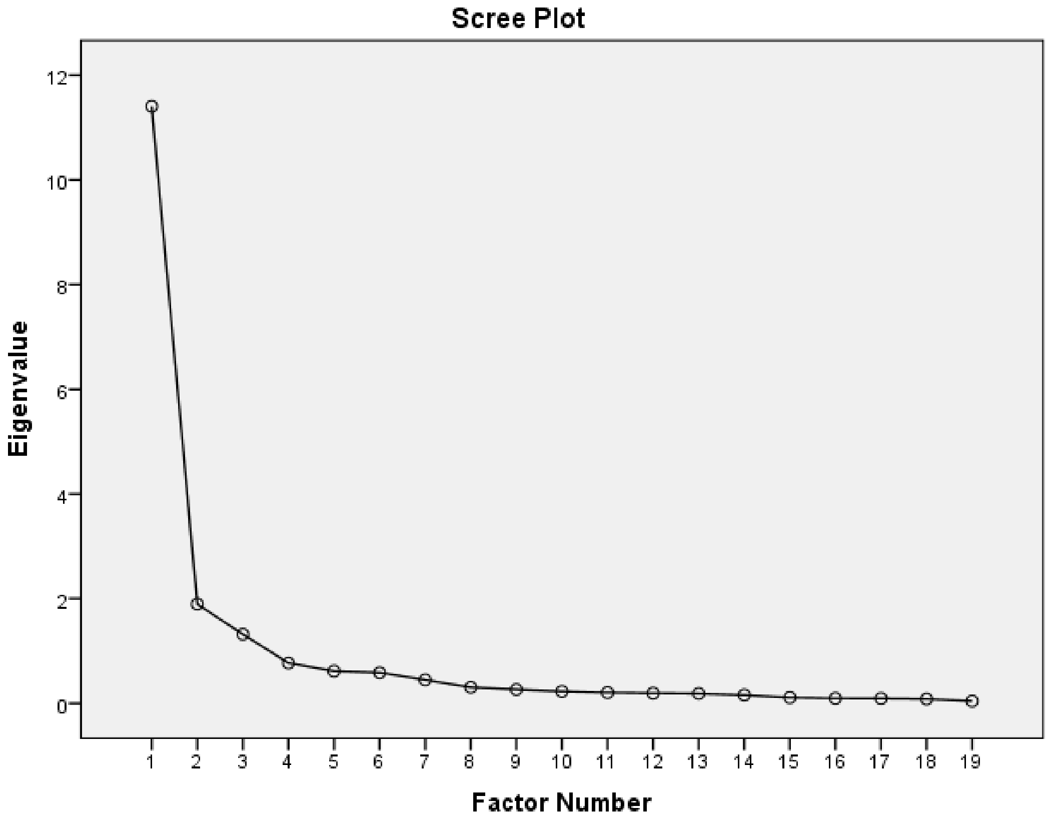
Joyfulness

+ keyed	Love life.
Radiate joy.
Feel lucky most of the time.
Just know that I will be a success.

– keyed	Am often in a bad mood.
Feel that my life lacks direction.
Have a dark outlook on the future.

Appendix 4: Scree plots and parallel analyses described in Chapter 5.4.2.3

Online sample



Run MATRIX procedure:

PARALLEL ANALYSIS:

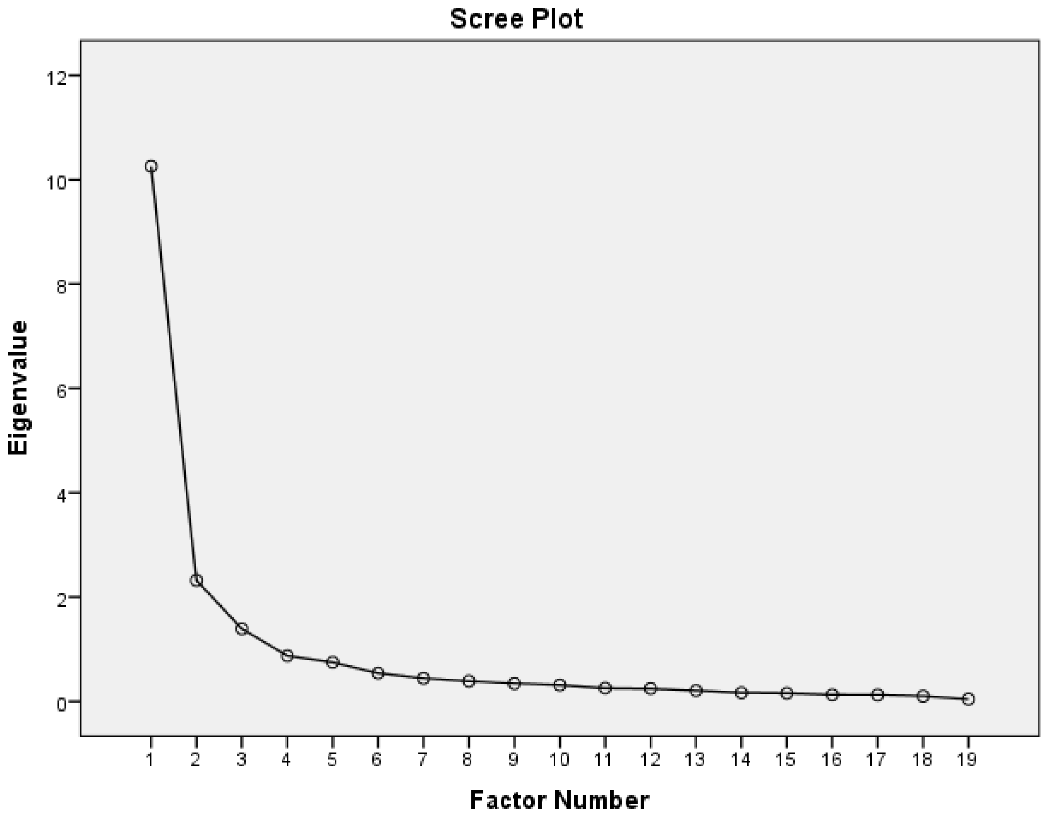
PAF/Common Factor Analysis & Random Normal Data Generation

Specifications for this Run:
Ncases 362
Nvars 19
Ndatsets 1000
Percent 95

Raw Data Eigenvalues, & Mean & Percentile Random Data Eigenvalues

Root	Raw Data	Means	Prcntyle
1.000000	11.205269	.485388	.573080
2.000000	1.678608	.401118	.468285
3.000000	1.045945	.336322	.387885
4.000000	.539301	.280445	.327487
5.000000	.434368	.232578	.275772
6.000000	.221620	.187939	.228626
7.000000	.114619	.146321	.181731
8.000000	.072878	.106341	.143398
9.000000	.060785	.068439	.104218
10.000000	.027318	.030770	.062961
11.000000	.005453	-.005333	.024230
12.000000	-.029570	-.039903	-.010280
13.000000	-.038538	-.075136	-.048549
14.000000	-.051930	-.109479	-.082053
15.000000	-.074172	-.144583	-.117580
16.000000	-.081919	-.179228	-.151759
17.000000	-.083709	-.215781	-.187049
18.000000	-.094198	-.254855	-.224984
19.000000	-.142206	-.301274	-.265659

British sample



Run MATRIX procedure:

PARALLEL ANALYSIS:

PAF/Common Factor Analysis & Random Normal Data Generation

Specifications for this Run:
Ncases 241
Nvars 19
Ndatsets 1000
Percent 95

Raw Data Eigenvalues, & Mean & Percentile Random Data Eigenvalues

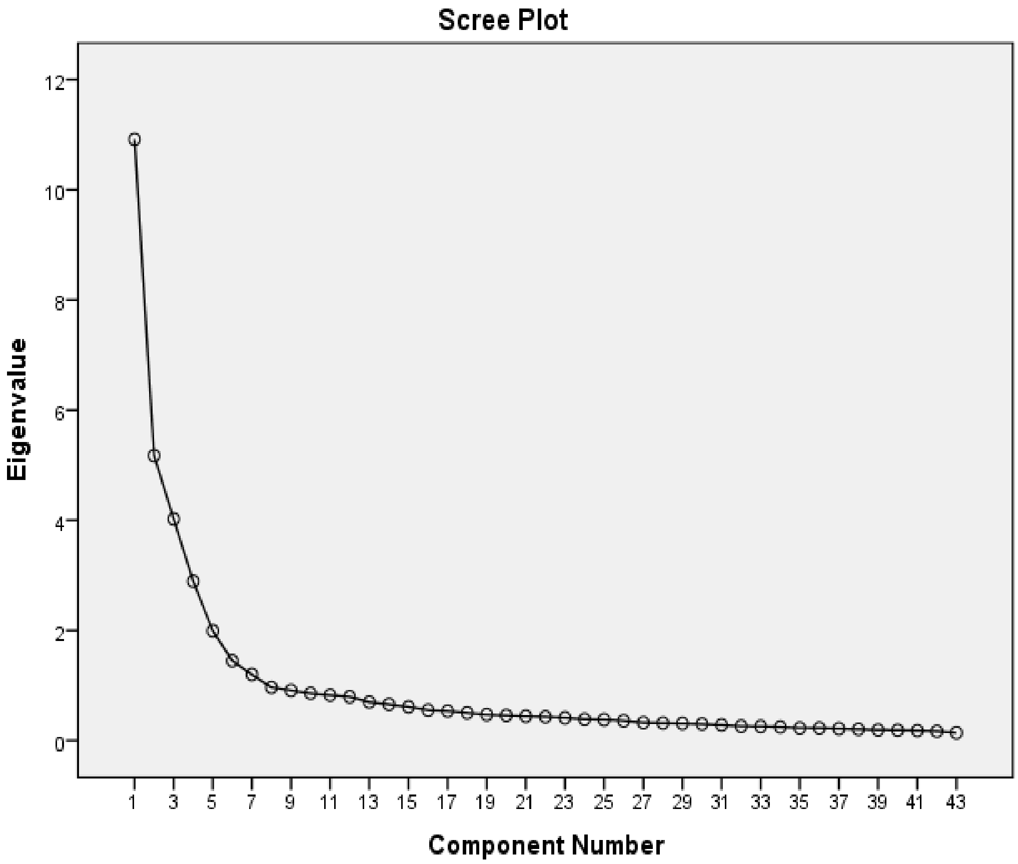
Root	Raw Data	Means	Prcntyle
1.000000	10.014638	.618141	.720712
2.000000	2.090454	.509936	.588936
3.000000	1.098899	.430692	.498115
4.000000	.574831	.361164	.419509
5.000000	.487616	.298523	.356533
6.000000	.261296	.242371	.293350
7.000000	.134928	.189428	.238249
8.000000	.072136	.138902	.184070
9.000000	.052842	.090530	.134336
10.000000	.007982	.045258	.086409
11.000000	-.030027	.001614	.037633
12.000000	-.033742	-.041261	-.007291
13.000000	-.040344	-.082314	-.049413
14.000000	-.053105	-.122893	-.090432
15.000000	-.096145	-.163661	-.132457
16.000000	-.103544	-.204995	-.172671
17.000000	-.108040	-.247177	-.216563
18.000000	-.117942	-.291579	-.259612
19.000000	-.129644	-.343817	-.303799

Appendix 5: Drive: Short (Drive:S; see Chapter 5.5.1.2)

Instructions: Below you find a list of characteristics describing people. Please use the rating scale below to describe how much/little of each characteristic you show or possess. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of roughly your age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence.

Item	1 Very Little	2	3	4	5	6	7 Very Much
1. Self-Confidence	❍	❍	❍	❍	❍	❍	❍
2. Zest/Enthusiasm/Vitality	❍	❍	❍	❍	❍	❍	❍
3. Valor/Bravery/Courage	❍	❍	❍	❍	❍	❍	❍
4. Liveliness	❍	❍	❍	❍	❍	❍	❍
5. Joyfulness	❍	❍	❍	❍	❍	❍	❍
6. Insight	❍	❍	❍	❍	❍	❍	❍
7. Initiative	❍	❍	❍	❍	❍	❍	❍
8. Diligence	❍	❍	❍	❍	❍	❍	❍
9. Generates Ideas	❍	❍	❍	❍	❍	❍	❍
10. Industriousness/Perseverance/Persistence	❍	❍	❍	❍	❍	❍	❍
11. Self-Discipline	❍	❍	❍	❍	❍	❍	❍
12. Achievement-Striving	❍	❍	❍	❍	❍	❍	❍
13. Hope/Optimism	❍	❍	❍	❍	❍	❍	❍

Appendix 6: Scree plot and parallel analysis described in Chapter 5.6.2.1



Run MATRIX procedure:

PARALLEL ANALYSIS:

Principal Components & Random Normal Data Generation

Specifications for this Run:
Ncases 475
Nvars 43
Ndatsets 1000
Percent 95

Raw Data Eigenvalues, & Mean & Percentile Random Data Eigenvalues

Root	Raw Data	Means	Prcntyle
1.000000	10.916084	1.625787	1.691612
2.000000	5.175818	1.558346	1.607106
3.000000	4.024936	1.506822	1.546229
4.000000	2.892673	1.462537	1.503051
5.000000	1.993271	1.423688	1.458108
6.000000	1.449236	1.386828	1.418604
7.000000	1.201156	1.353798	1.383650
8.000000	.967247	1.321949	1.350122
9.000000	.910298	1.291609	1.318759
10.000000	.857949	1.262745	1.288942
11.000000	.826088	1.234511	1.259609
12.000000	.795297	1.207547	1.232277
13.000000	.699300	1.181310	1.205877
14.000000	.658868	1.155799	1.179753
15.000000	.613130	1.131330	1.155292
16.000000	.553651	1.107150	1.130285
17.000000	.537033	1.083005	1.105654
18.000000	.504252	1.059918	1.081637
19.000000	.471558	1.037026	1.058826
20.000000	.457320	1.015224	1.035731
21.000000	.443133	.993572	1.015518
22.000000	.431975	.971053	.990980
23.000000	.413926	.949826	.970247
24.000000	.388744	.929023	.949669
25.000000	.381940	.908765	.929101
26.000000	.361597	.888463	.908530
27.000000	.326496	.867943	.888098
28.000000	.318417	.847431	.866141
29.000000	.311156	.827659	.847450
30.000000	.301771	.807578	.827170
31.000000	.284286	.787238	.807577
32.000000	.263027	.767324	.788647
33.000000	.257859	.747836	.767332
34.000000	.247981	.727752	.747032
35.000000	.228916	.707909	.727387
36.000000	.227807	.686978	.707194
37.000000	.218037	.666472	.688070
38.000000	.206009	.644808	.667962
39.000000	.195295	.623592	.646497
40.000000	.190130	.600743	.625510
41.000000	.182712	.576505	.602590
42.000000	.169959	.549257	.573805
43.000000	.143661	.515347	.545210

Appendix 7: Related papers

Chapter 2

Siegling, A. B., Petrides, K. V., & Martskvishvili, K. (2015). An examination of a new psychometric method for optimising multi-faceted assessment instruments in the context of trait emotional intelligence. *European Journal of Personality*, *29*(1), 42–54. doi:10.1002/per.1976

Chapter 3

Study 1

Siegling, A. B., & Petrides, K. V. (2014). Measures of trait mindfulness: Convergent validity, shared dimensionality, and linkages to the Five-Factor Model. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.01164

Study 2

Siegling, A. B., & Petrides, K. V. (accepted). Zeroing in on mindfulness facets: Similarities, validity, and dimensionality across three independent measures. *PLOSOne*.

Chapter 4

Siegling, A. B., Furnham, A., & Petrides, K. V. (under review). Facet Benchmarking: A rigorous application of a new psychometric method for refining multi-faceted assessment instruments. *European Journal of Personality*.

Chapter 5

Siegling, A. B., & Petrides, K. V. (under review). Development and validation of a measure of general motivation. *Journal of Organizational Behavior*.

Chapter 6

Siegling, A. B., & Petrides, K. V. (in preparation). Facet Benchmarking: Application of a psychometric method for refining multi-faceted assessment instruments to a new measure of motivation.