The Pennsylvania State University

The Graduate School

College of the Liberal Arts

FACTOR STRUCTURE, MEASUREMENT EQUIVALENCE, AND CRITERION RELATIONS OF THE INVENTORY OF PERSONALITY ORGANIZATION (IPO) IN TWO NONCLINICAL SAMPLES

A Thesis in

Psychology

by

William D. Ellison

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2009

The thesis of William D. Ellison was reviewed and approved* by the following:

Kenneth N. Levy
Assistant Professor of Psychology
Thesis Advisor

Aaron Pincus
Associate Professor of Psychology

Reginald Adams
Assistant Professor of Psychology

Melvin M. Mark
Professor of Psychology
Head of the Department of Psychology

*Signatures are on file in the Graduate School.

ABSTRACT

Borderline Personality Disorder (BPD) represents a serious public health risk. Individuals with BPD consume a disproportionately high amount of mental health services, engage in frequent suicidal and parasuicidal behavior, and do not respond as well to established treatments for other disorders. Research into BPD depends on the availability of well-validated and useful measures of its core features. However, few such instruments exist. The present report examines the psychometric performance of the Inventory of Personality Organization (IPO), a theory-based self-report measure of borderline personality features consisting of 46 items. The IPO was administered to a total of 2775 college undergraduates from two separate universities in the northeastern U.S. A confirmatory factor analysis suggested that the IPO is better characterized by a three-factor than a two-factor model, which is consistent with its theoretical structure. A formal test of measurement equivalence suggested that the IPO generalizes well to different populations. The relationship between IPO subscales and measures of identity, defense mechanisms, affect, and reckless and self-harming behavior was also examined. Results indicated that the identity diffusion subscale of the IPO displayed good convergent and discriminant validity.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1. INTRODUCTION

Borderline Personality Disorder (BPD) represents a serious public health risk. Studies typically estimate its prevalence at around 1-3% of the general population (Lenzenweger, Lane, Loranger, & Kessler, 2007; Samuels et al., 2002; Torgersen, Kringlen, & Cramer, 2001), although a recent major epidemiological study (Grant et al., 2008) suggests that its prevalence may be as high as 6%. Individuals with BPD consume a large amount of mental health services, making up over 9% of psychiatric outpatients (Zimmerman, Rothschild, & Chelminski, 2005) and 15-20% of psychiatric inpatients (Widiger & Frances, 1989). In fact, individuals with BPD use more mental health services of every type than those with major depression (Bender et al., 2001). Around 70% of individuals with BPD commit repeated self-injurious acts (Clarkin, Widiger, Frances, Hurt, & Gilmore, 1983), and up to 10% eventually commit suicide (Stone, 1993). BPD is substantially comorbid with other mental disorders (Nurnburg et al., 1991; Skodol et al., 2002; Zanarini et al., 1998; Zanarini et al., 2004), and there is evidence that the presence of BPD negatively affects outcome in the treatment of depression (Shea, Widiger, & Klein, 1992), anxiety disorders (Chambless, Renneberg, Goldstein, & Gracely, 1992; Cloitre & Koenen, 2001; Mennin & Heimberg, 2000), and eating disorders (Cooper, Coker, & Fleming, 1996). Although several promising psychosocial treatments have recently been developed for BPD itself (Bateman & Fonagy, 1999; Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Giesen-Bloo et al., 2006; Linehan, 1993), therapy with this population remains a challenge. Because of its chronic course, high lethality, extensive comorbidity, and negative impact on treatment, it is especially important that the mental health community understand BPD.

Conceptualizations of BPD are currently dominated by the model presented in the fourth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental*

Disorders (DSM; American Psychiatric Association, 1994). This model conceives of the disorder as a categorical and unitary disease entity, the presence or absence of which in a particular individual depends on the presence or absence of nine criteria, or symptoms. These symptoms include frantic efforts to avoid abandonment, unstable interpersonal relationships, identity disturbance, impulsivity, suicidal or self-injurious behavior, affective instability, chronic feelings of emptiness, inappropriate anger, and transient, stress-related paranoia or dissociation (American Psychiatric Association, 1994). Five or more of these symptoms are needed in order to qualify for the diagnosis. Importantly, the DSM makes no theoretical claims about which symptoms might be fundamental in BPD. Each criterion is considered an equally informative sign that the disorder is (or is not) present.

Because of the prominence of the DSM diagnostic system in contemporary mental health practice and research, most methods of assessing Borderline Personality Disorder adopt a similar symptom-based, atheoretical stance. For example, clinical interviews such as the International Personality Disorder Examination (IPDE; Loranger, 1999) and self-report measures such as the Schedule for Nonadaptive and Adaptive Personality (SNAP; Clark, 1993) assess for the presence of the DSM symptoms of BPD. After a decision has been reached about whether each symptom is present, the number of symptoms is then totaled to determine whether an individual qualifies for the diagnosis.

However, although the DSM strives to be theory-neutral, several observers (e.g., Blashfield & Livesley, 1991; Carson, 1991; Millon, 1991) have pointed out the impossibility of this standard. These authors all note that the adoption of a symptom-based, categorical system of diagnosis was itself implicitly based on a medical theory of disorder. Moreover, because the diagnostic criteria were developed by committee in an attempt to be useful to clinicians and researchers of different backgrounds (Spitzer, 2001), in many cases the criteria for a given disorder are a hodgepodge of different theoretical perspectives instead of a single objective list. As a result, many observers have criticized the construct validity of the resultant categories. Indeed, Blatt and Levy (1998) argue that an atheoretical diagnostic system is not only impossible, but also undesirable, because it largely fails to specify the theoretical relationships researchers would need to test in order to validate its categories. As Carson (1991) notes, the set of criteria that makes up a DSM diagnosis "conveys almost nothing about the presumed entity that can be described without referring back to these same terms or their close synonyms" (p. 304). Because the DSM remains silent on how the markers of BPD might relate to one another, as well as how they might relate to the etiology and correlates of the disorder, researchers interested in such questions must rely on outside theories to guide their efforts.

One of the most prominent theories delineating the relationships between different characteristics of borderline pathology is Otto Kernberg's (1975) object relations theory of borderline personality organization (BPO). Kernberg (1975; Clarkin et al., 2007) describes three basic levels of personality organization: neurotic, borderline, and psychotic. The borderline personality organization category describes not only individuals with a DSM diagnosis of BPD, but also those with other severe personality disorders with similar psychological characteristics (Kernberg, 1996).

According to Kernberg (1996), three structural qualities distinguish between the three levels of personality organization. Primary among these is the concept of identity diffusion, or the inability to integrate positive and negative units of self-knowledge, which results in an incoherent sense of self that is distressing to the individual. Another key factor in Kernberg's theory of personality organization is the developmental maturity of defense mechanisms typically employed. In particular, Kernberg theorizes that the use of primitive (that is, immature) defenses, as opposed to more adaptive mechanisms, is important in determining the individual's level of personality organization. The final variable of importance is reality testing, or "the capacity to differentiate self from non-self, intrapsychic from external stimuli, and to maintain empathy with ordinary social criteria of reality" (Kernberg, 1996, p. 120). In Kernberg's analysis, individuals at the neurotic level of personality organization have a well-consolidated sense of self without identity diffusion, use mature instead of primitive defenses, and have intact reality testing. Those at the borderline level have identity diffusion and use primitive defenses but largely maintain the capacity to test reality. Finally, the psychotic level of organization is characterized by loss of reality testing as well as identity diffusion and primitive defenses (Table 1).

Table 1

Structural Summary of Psychotic, Borderline, and Neurotic Levels of Personality Organization

Personality Organization	Identity Diffusion	Primitive Defenses	Poor Reality Testing
Psychotic	+	+	+
Borderline	+	+	–
Neurotic	–	–	–

Source: Oldham et al. (1985).

The assessment of personality organization according to this theory is traditionally accomplished through a clinical interview. However, in recent years, a number of alternative methods have been developed, including a structured interview (Clarkin, Caligor, Stern, & Kernberg, 2002), a clinician-rated instrument (Hébert et al., 2003), and self-report measures (e.g., Kernberg & Clarkin, 1995). Of these four methods of assessment, self-report questionnaires are undoubtedly the easiest to administer and interpret, meaning that they can be administered to large samples in research studies. For this reason, self-report measures assume a primary importance in research aimed at validating Kernberg's theory of personality organization. This report will address the psychometric properties of such a questionnaire.

As early as 1985, John Clarkin and colleagues (Oldham et al., 1985) made efforts to develop a self-report measure relating to Kernberg's theory of personality organization. This preliminary scale was divided into three subscales: identity diffusion, reality testing, and primitive defenses. Questions were rated on a 5-point scale from "never true" to "always true." 138 items were tested, and the top 10 items were selected for each subscale based on the correlation between the item and the scale score, resulting in a 30-item scale. The subscales showed high internal consistencies (Cronbach's α values ranging from 0.84 to 0.92) and high levels of intercorrelation. Pearson's r coefficients between the scales were 0.76 for identity diffusion and primitive defenses, 0.79 between identity diffusion and reality testing, and 0.75 between primitive defenses and reality testing.

A later version of this scale was developed and named the Inventory of Personality Organization (IPO; Kernberg & Clarkin, 1995). This new instrument added several items to each primary scale and also included several other subscales corresponding to other pathological personality styles, moral values, aggression, and quality of object relations. Lenzenweger, Clarkin, Kernberg, & Foelsch (2001) found that the three primary subscales had high internal consistencies (Cronbach's α's ranging from 0.81 to 0.88) and, again, high intercorrelations. Primitive defenses and identity diffusion were highly related (r = 0.82 in Study 1, r = 0.83 in Study 2), and the other scales displayed moderate to high intercorrelations (primitive defenses with reality testing, r = 0.65 and r = 0.76; identity diffusion with reality testing, r = 0.62 and r = 0.73).

This pattern of intercorrelations among IPO subscales was also noted by Normandin et al. (2002) in a study of a French version of the IPO. Normandin and colleagues conducted a confirmatory factor analysis and found an especially high correlation between factors composed of primitive defenses and identity diffusion items (r = 0.69). In contrast, primitive defenses and reality testing displayed only a moderate correlation (r = 0.49), and identity diffusion and reality testing were weakly related (r = 0.27).

It is entirely consistent with Kernberg's theory of personality organization that identity diffusion and primitive defenses subscales would correlate particularly strongly. Kernberg (1975) theorizes that a particular immature defense mechanism, called splitting, contributes to the development of identity diffusion. Splitting involves the motivated separation of positive and negative representations of self and others in an effort to protect the positive qualities from contamination. According to Kernberg, this type of representational structure characterizes the object relations of young children, and an important task of mature development is the integration of positive and negative representations into more realistic (that is, mixed) models of self and other. It is the failure to accomplish this integration that produces the vague, incoherent sense of self observed in borderline individuals. Moreover, the identity diffusion that results from a reliance on splitting prevents the individual from being able to achieve integration by denying him or her access to thoughts that would neutralize extreme emotional impulses. Thus, primitive defenses (especially splitting) and identity diffusion form a "vicious circle" (Kernberg, 1975, p. 29) in Kernberg's theory of personality organization. Because of this intimate functional relationship and the fact that identity diffusion and primitive defenses always covary across levels of personality organization (Table 1), it is not surprising that scales based on these constructs are closely related.

However, for the purposes of assessment, it is difficult to justify the separation of identity diffusion and primitive defenses items into two subscales. Although the two constructs may be theoretically divergent and psychologically distinct, the self-report subscales based on them may be psychometrically redundant. Concretely, if the correlations between the scales begin to approach the inter-item correlations within the scales, it makes little sense to divide the items into two groups (Clark & Watson, 1995). It will be especially important, therefore, to demonstrate that the primitive defenses and identity diffusion subscales of the IPO have discriminant validity, given the data cited above.

A recent paper by Lenzenweger, Clarkin, Kernberg, & Foelsch (2001) attempted to investigate this issue. The authors conducted a confirmatory factor analysis of the 1995 version of the IPO and tested a three-factor model (with factors corresponding to identity diffusion, primitive defenses, and reality testing) against a two-factor model (with identity diffusion and primitive defenses items loading on a single factor). They found that the two-factor model fit roughly as well as the three-factor model and concluded that the two-factor structure provided the best mix of fit and parsimony. They also conducted tests of criterion-related validity and found relationship patterns with affect, aggression, schizotypy, self-consciousness, and self-monitoring that they concluded were consistent with Kernberg's (1975) model of borderline personality organization.

However, the analysis presented by Lenzenweger et al. (2001) has several weaknesses that limit the usefulness of the authors' conclusions. First, the confirmatory factor analysis was conducted on a sample of only 249 participants. Because of the fairly large model being estimated (57 observed indicators loading on two or three intercorrelated latent factors), a large sample size is needed in order to provide stable estimates of the model's parameters (Brown, 2006). Although there are no definite guidelines specifying the sample size needed for such a model, a simulation study by Dolan (1994) showed that, even under ideal conditions (a model of only 8 indicators with a normally-distributed response pattern), a sample size of 200 produced a very large bias in both standard errors and parameter estimates. This bias is likely to increase with substantially larger models and non-normal data (Sharma, Durvasula, & Dillon, 1989), both of which characterize Lenzenweger and colleagues' analysis.

The authors attempt to correct for the non-normal distribution of their data by using three different estimation procedures: maximum likelihood (ML) estimation based on normal theory, ML estimation with corrected chi-square statistics, and iteratively reweighted generalized least squares estimation (Lenzenweger et al., 2001). However, each of these estimations is statistically equivalent, as iteratively reweighted generalized least squares estimation has been shown to converge asymptotically to maximum likelihood estimation (Goldstein, 1986). The authors, therefore, used three procedures with identical underlying estimation methods. The fact that these procedures produced identical results may be a reflection of common bias, not an indication that they converge on a common and generalizable solution. The results of Lenzenweger and colleagues' factor analysis should thus be considered tentative.

The composition of the sample in Lenzenweger et al. (2001) also constrains the usefulness of their factor analysis. The participants in the study were undergraduates from Cornell University, an elite private college. Although the authors note that undergraduates at Cornell have been found to be “generally representative of other young people their age in the community on many other psychological and psychopathological dimensions” (Lenzenweger et al., p. 579), samples from the same research group, drawn from the same student population, are elsewhere noted to differ from the population at large in ethnicity and socioeconomic status (Lenzenweger, 1999). It is also likely that Cornell undergraduates differ substantially from even other college students in academic achievement, general level of functioning, and other variables that may relate to Borderline Personality Disorder. To the extent that these variables matter in the expression of borderline personality organization, the composition of the sample may be expected to alter the psychometric performance of the IPO.

The current study, like that of Lenzenweger et al. (2001), is aimed at evaluating whether identity diffusion and primitive defenses should be conceptualized as separate subscales or whether they should be collapsed for assessment purposes. In order to do this, the factor structure of the newest version of the IPO (Clarkin, Foelsch, & Kernberg, 2001), consisting of 46 items delineating putative identity diffusion, primitive defenses, and reality testing subscales, was examined in two non-clinical samples of young adults. This version of the IPO was developed by eliminating items from the 57-item version that contributed negatively to the internal consistency (Cronbach’s α) values of the three clinical scales. Thus, the new version may show a distinct latent structure from previous editions, but its psychometric properties are as yet unknown. The object of the current paper was thus to conduct an initial inquiry into whether a three-factor model characterizes the 46-item IPO or whether its identity diffusion and primitive defenses subscales are largely indistinct, as Lenzenweger and colleagues concluded about the 57-item measure.

The current study was designed to avoid several of the limitations of Lenzenweger et al.'s (2001) analysis of the 57-item IPO. First, each sample is quite large, roughly four times as large as the one used in Lenzenweger et al. This advantage will enhance confidence in the stability of the obtained parameter estimates. In addition, it will allow for the use of different estimation procedures (e.g., weighted least squares estimation) that may be more appropriate for non-normal data than the methods used by Lenzenweger and colleagues.

Second, the composition of the current samples differs markedly from the sample of Cornell undergraduates used by Lenzenweger et al. (2001) in several potentially important ways, such as socioeconomic status and ethnicity. The present investigation thus provides an opportunity to draw conclusions about the generalizability of the psychometric properties of the 46-item IPO in different samples. Moreover, the use of two samples in the current study, each of which differs from the other in several domains, allows for an explicit test of the measurement equivalence of the IPO across these two groups.

Third, whereas Lenzenweger et al. (2001) used criterion measures that only partially overlapped with identity diffusion (the Self-Monitoring Scale, Gangestad and Snyder, 1985, and the Self-Consciousness Scale, Fenigstein et al., 1975) and did not provide evidence for the validity of the primitive defenses subscale, the current study relates these subscales of the 46-item IPO to several questionnaire measures of self-organization and defense styles. In particular, measures developed by social and personality psychologists whose items suggest a strong theoretical link to identity diffusion are used in order to examine the proposition, put forth by Lenzenweger and colleagues, that identity diffusion represents a uniquely pathological construct.

Instruments designed to assess defense mechanisms are also used in order to validate the primitive defenses subscale of the IPO. In addition, other constructs related to Borderline Personality Disorder are measured (e.g., deliberate self-harm, reckless behavior, and affect) in order to determine whether the IPO can be used to predict other markers of BPD among young adults. Importantly, these criterion measures are used to examine the possibility that the identity diffusion and primitive defenses subscales of the 46-item IPO might show divergent relationships with external constructs, even if they display a strong intercorrelation.

Chapter 2. METHODS

Data Collection

Two samples were used to evaluate the psychometric properties of the IPO. Table 2 compares characteristics of the current samples with the sample used in the confirmatory factor analysis conducted by Lenzenweger et al. (2001).

Sample 1

1381 undergraduates from Hunter College in New York City completed the IPO as part of a class requirement. This sample was similar in age (*M* = 20.8, *SD* = 4.6) to the sample in Lenzenweger et al. (2001), but there were also notable differences. A larger proportion of the Hunter College sample was female (74.5%), and the sample was ethnically quite diverse. Only 30.1% of the Hunter College sample was Caucasian, whereas 12.1% was African-American, 19.8% Hispanic, and 19.5% East Asian or South Asian. 18.5% listed their primary ethnicity as "other." Participants were recruited from an undergraduate subject pool comprised of students in introductory psychology classes. The questionnaire battery was administered in groups of 20-30 participants.

Measures

Inventory of Personality Organization. The Inventory of Personality Organization (IPO) is a self-report questionnaire designed to measure constructs related to Kernberg's (1975) theory of borderline personality organization. Whereas Lenzenweger et al. (2001) used a preliminary version (Kernberg & Clarkin, 1995) in their factor analysis, the current study uses a more recent version (Clarkin, Foelsch, & Kernberg, 2001). There are three primary clinical subscales of the IPO, which were derived on the basis of clinical theory and comprise 46 items. These are

Table 2

Demographic Characteristics of Current Samples and Sample in Lenzenweger et al. (2001)

	Current samples		Lenzenweger et al. (2001)
	Hunter College	**Penn State**	
Sample size	**1381**	**1394**	249
Mean age	**20.83**	**19.05**	19.67
% female	**74.5**	**66.2**	65.5
Primary race/ethnicity (%)			
African-American	**12.1**	**3.4**	4.4
Latino/Hispanic	**19.8**	**2.2**	4.8
Caucasian	**30.1**	**84.1**	70.3
East/South Asian	**19.5**	**5.4**	17.7[a]
Other	**18.5**	**4.0**	2.8

[a] Asian-Pacific Islander

identity diffusion (14 items), or the extent to which the individual's self-concept is clearly defined; primitive defenses (19 items), or the extent to which the individual uses psychological defenses that are thought to reflect an immature or primitive style; and reality testing (13 items), or the extent to which the individual is able to maintain empathy with socially appropriate standards of reality. Items are rated on a 5-point Likert-type scale ranging from "never true" to "always true." Sample items include "People tell me I behave in contradictory ways" (from the primitive defenses subscale), "My life goals change frequently from year to year" (identity diffusion), and “I have seen things which do not exist in reality” (reality testing). Reliability and validity data for earlier versions of the IPO are discussed above.

Defenses. In order to establish the criterion validity of the primitive defenses subscale of the IPO, several questionnaire measures of defense mechanisms were administered. These included the Defense Style Questionnaire (DSQ-40; Andrews, Singh, & Bond, 1993), which uses 40 items rated on a 9-point Likert-type scale to measure 20 distinct defense mechanisms. The measure has shown adequate internal consistency and reliability (Andrews et al., 1993). Its latent structure varies from study to study (Andrews et al., 1993; Ruuttu et al., 2006; Trijsburg, Van T’Spijker, Van, Hesselink, & Duivenvoorden, 2000), but the theoretically-based hierarchy of defense styles predicts severity of psychiatric symptoms (Watson, 2002; Ruuttu et al., 2006) and discriminates between individuals with Borderline Personality Disorder and other personality disorders (Bond, Paris, & Zweig-Frank, 1994). The DSQ has also been shown to discriminate between psychiatric outpatients and healthy controls (Trijsburg et al., 2000). In the current study, the “immature” defenses identified by Andrews, Singh, and Bond (1993) were used to test the convergent validity of the primitive defenses subscale of the IPO. These 12 defenses were collapsed into a single index comprised of 24 items in order to produce a measure of these defenses as a whole and to facilitate the interpretation of their relationship to the IPO.

In order to explore the relationship of the primitive defenses subscale with the specific defense of splitting, which Kernberg (1975) calls “an essential defensive operation of the borderline personality organization” (p. 29), the Splitting Scale (Gerson, 1984) was included. This scale was designed to synthesize Kernberg’s (1975) and Kohut’s (1968) conceptualization of splitting. It includes items relating to different emotions that might be the target of splitting, such as anger, as well as items reflecting the idealization, grandiosity, and exhibitionism that may theoretically result from this defensive operation. In addition, one item was included to capture identity diffusion, which Gerson understood as a consequence of splitting. Items on this 14-item self-report measure are rated on a 7-point Likert-type scale ranging from "not at all true" to "very true." Sample items include "It's hard for me to get angry at people I like," and "When I'm angry, everyone around me seems rotten." A factor analysis (Gerson, 1984) suggested that one principal factor, composed of 10 of the 14 items, accounted for over 45% of the variance in the measure.

Affect. The primary measure of affect in this sample is the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). The PANAS is a 20-item self-report questionnaire designed to measure positive and negative affect, with 10 items in each subscale. Participants were asked to rate on a 5-point Likert-type scale to what extent they had been feeling certain emotions in the previous few months, ranging from "Very slightly or not at all" to "Extremely." This form of the PANAS has been extensively validated in diverse samples (Crawford & Henry, 2004; Huebner & Dew, 1995; Watson et al.) and shows adequate internal and test-retest reliability (Watson et al.).

Sample 2

1394 undergraduates at Pennsylvania State University in State College, Pennsylvania completed a questionnaire battery including the IPO as part of a research requirement for an introductory psychology course. Data collection occurred via an Internet survey on the Surveymonkey data collection website (http://www.surveymonkey.com). In contrast to the Hunter College sample, and similar to the sample in Lenzenweger et al. (2001), the Penn State sample was predominantly Caucasian (see Table 2). Likewise, the sample had a similar mean age and proportion of female respondents as the sample used by Lenzenweger et al.

Surveys were given to the Penn State sample over the Internet, leading to the risk that some participants may have filled out questionnaires thoughtlessly or under distraction. Therefore, responses on the Jackson Infrequency Scale (Jackson, 1970) were used to purge the data of respondents who gave random or stereotyped answers, such that data from participants who answered more than two items on the Jackson Scale in the infrequent direction were not used in analyses.

Measures

Inventory of Personality Organization. The same version of the IPO was administered to this sample as was administered to the Hunter College sample (see above).

Self-concept. In order to investigate the similarity of the IPO identity diffusion subscale to other self-report measures of self-concept structure, several other questionnaires related to the self were administered. These included the Self-Concept Clarity Scale (SCCS; Campbell et al., 1996), a 12-item measure designed to assess the extent to which a person's self-beliefs are clearly and confidently defined, internally consistent, and stable. Items are rated on a 5-point Likert-type scale ranging from “Strongly disagree” to “Strongly agree.” A sample item is “my beliefs about myself often conflict with one another.” The SCCS has shown high internal consistency (Cronbach’s $\alpha = .86$) and predicts the actual consistency of individuals’ self-attribute ratings (Campbell et al.). Self-concept clarity has also been shown to relate to various measures of psychological adjustment (Campbell, Assanand, & Di Paula, 2003).

Participants also completed the Borderline Personality Inventory (BPI; Leichsenring, 1999), a 53-item self-report measure designed to assess borderline personality organization. The BPI is explicitly based on Kernberg’s model of personality organization (though developed independently) and, like the IPO, has three subscales corresponding to identity diffusion, primitive defenses, and reality testing. Unlike the IPO, item responses are made on a true-false basis. An example item from the identity diffusion subscale is “Sometimes I act or feel in a way that does not fit me.” Internal consistency of the three primary subscales was mixed in Leichsenring’s preliminary study; identity diffusion (Cronbach’s $\alpha = .83$) and primitive defenses ($\alpha = .81$) had high consistency, but reality testing was somewhat lower ($\alpha = .68$). There is also some indication that the scale has a six-factor structure, at least in adolescents (Chabrol et al., 2004). However, the BPI subscales discriminate well between inpatients with Borderline Personality Disorder, inpatients without BPD, inpatients with schizophrenia, and non-inpatient controls (Leichsenring, 1999).

A subset of 801 participants completed Ziller’s Self-Complexity Measure (SCM; Ziller, Martell, & Morrison, 1977). This measure consists of 109 trait adjectives (e.g., clever, patient) presented to the participant along with the instructions, “Please check each adjective on the list below that you think describes you.” Self-complexity is measured by the number of adjectives freely checked. Ziller et al. (1977) found that self-complexity related to self-reported complexity of self-concept and other-rated complexity of self-portrait photographs, but not to intelligence.

The same subset of 801 participants completed a modified version of Rosenberg’s (1965) Stability of Self Scale (SSS), which consists of 5 items measuring the self-reported stability of self-esteem. In the current study, Marsh’s (1993) modified version of this scale was used, in which items were rated on a 5-point Likert-type scale ranging from “Never true” to “Almost always true.” Marsh found that scores on the SSS were unrelated to actual intraindividual variability in self-esteem, suggesting that the measure does not perform as designed. However, the items (e.g., “My opinion of myself tends to change a good deal instead of always remaining the same”) are extremely similar to self-report measures of self-concept stability, such as the

IPO. Therefore, the SSS is included, in part, in order to investigate whether the IPO is able to discriminate between self-reported self-esteem stability and self-reported self-concept stability.

A smaller subset of 303 participants completed the Life Problems Inventory (LPI; Rathus & Miller, 1995), a 60-item self-report scale. One subscale, called “Confusion about Self” by the authors, contains items such as “I’m not sure I know who I am or what I want in life.” Responses are made on a 5-point Likert-type scale ranging from “Not at all like me” to “Extremely like me.” To date, no studies have examined the psychometric properties of this inventory or its subscales.

The same subset of 303 participants was given the Differentiation of Self Inventory (DSI; Skowron & Friedlander, 1998). This 43-item self-report measure was designed to evaluate individuals’ differentiation of emotional and intellectual aspects of self and external situations, as well as individuals’ ability to balance intimacy and autonomy. Items are rated on a 6-point Likert-type scale ranging from “Not at all true of me” to “Very true of me.” An example item is “No matter what happens in my life, I know that I’ll never lose my sense of who I am.” Initial studies have demonstrated the high internal consistency of the DSI (Skowron & Friedlander) and revealed relationships between this scale and measures of emotional and interpersonal regulation (Skowron, Holmes, & Sabatelli, 2003; Skowron, Wester, & Azen, 2004; Wei, Vogel, Ku, & Zakalik, 2005).

Affect. In order to examine the relationship between IPO factors and affective phenomena typical of individuals with BPD, two measures pertaining to emotional reactivity and intensity were used. Participants in the Penn State sample completed the Affect Lability Scales (ALS; Harvey, Greenberg, & Serper, 1989), a 54-item instrument designed to measure lability in anxiety, depression, anger, and hypomania, and labile shifts between anxiety and depression and hypomania and depression. A sample item is “One minute I can be feeling O.K., and the next minute I’m tense, jittery, and nervous.” Items are rated on a 4-point Likert-type scale ranging from “Very characteristic of me, extremely descriptive” to “Very uncharacteristic of me, extremely undescriptive.” Initial research (Harvey et al., 1989) showed that the scales have adequate internal consistency and strong test-retest reliability and correlate with BDI scores but not with affect intensity. The individual scales are also highly intercorrelated, suggesting that they measure a general tendency toward emotional lability (Harvey et al.). Scores on the ALS have been found to differentiate between Borderline Personality Disorder and Bipolar II Disorder (Henry et al., 2001) and to predict lifetime incidence of anxiety and depression among personality-disordered individuals (Bunce & Coccaro, 1999). Because of a clerical error, the last three items were omitted from the ALS questionnaire given to participants. Two of the omitted items were components of the elation subscale, and the remaining item belonged to the depression subscale.

In addition, the Affect Intensity Measure (AIM; Larsen, Diener, & Emmons, 1986) was used to investigate whether the IPO might be associated with the intensity of emotion often noted in individuals with BPD (American Psychiatric Association, 1994). The AIM is a 40-item measure designed to assess intensity of an individual's affective responsiveness and contains items such as “When I accomplish something difficult I feel delighted or elated.” Items are rated on a 6-point Likert-type scale ranging from “Never” to “Always.” Several studies have shown that individuals with BPD show elevated scores on this measure as compared to normal or psychiatric controls (Flett & Hewitt, 1995; Henry et al., 2001; Levine, Marziali, & Hood, 1997; Yen, Zlotnick, & Costello, 2002), and the AIM also differentiates BPD from Bipolar II Disorder (Henry et al.).

Reckless and self-injurious behavior. Two measures of reckless and self-harming behavior, which is common in BPD (American Psychiatric Association, 1994), were included in order to investigate their relationship with IPO subscales. The first of these was the Deliberate Self-Harm Inventory (DSHI; Gratz, 2001), a 17-item self-report measure of diverse methods of self-injury that are typical of borderline individuals. The instrument measures both the lifetime use of each method (e.g., "Have you ever intentionally burned yourself with a cigarette?") and about its onset, frequency, and intensity. The DSHI shows adequate internal and test-retest reliability as well as theoretically appropriate criterion relations with measures of self-injury, suicidality, borderline personality organization, and other measures of psychopathology (Gratz, 2001; Fliege et al., 2006).

In addition, the Reckless Behavior Questionnaire (RBQ; Shaw, Wagner, Arnett, & Aber, 1991) was included in the questionnaire battery. The RBQ is a 10-item scale designed to measure reckless behavior in adolescents. Items pertain to the frequency, in the past year, of different types of potentially dangerous behavior (e.g., driving under the influence, drug use). Items are rated on a 5-point scale, with choices as follows: "0 times," "once," "2-5 times," "6-10 times," and "more than 10 times." Research using both high school and college samples (Shaw et al., 1991) supports a two-factor structure for the RBQ, where one factor consists of recklessness in driving behavior and vandalism and the other of recklessness in drug use and sex. A divergent pattern of criterion relations was found by gender. In males, both factors related strongly to measures of sensation seeking and aggression, whereas in females, the relationship of the drugs/sex factor to these criteria was much stronger than the driving/vandalism factor.

Statistical Analyses

Estimation procedure

Confirmatory factor analysis was carried out by means of the PRELIS and LISREL programs, version 7 (Jöreskog & Sorböm, 1989) and Mplus, version 5.1 (Muthén & Muthén, 2007). As Lenzenweger and colleagues (2001) noted, several properties of IPO data warrant close consideration. Items are rated on a 5-point Likert scale, suggesting that responses may not be properly considered continuous. Likewise, because the IPO assesses constructs that pertain to abnormal functioning, responses are likely to have heavily skewed distributions when they come from a non-clinical sample. Indeed, Lenzenweger and colleagues (2001) found a Mardia coefficient of 336.40 for the 57 items comprising their IPO subscales, indicating a substantial deviation from multivariate normality in their data. Because the current version of the IPO consists of a majority (46) of these items, multivariate normality was not assumed. The literature suggests that maximum likelihood estimation (for which normal data are assumed) may not be appropriate for non-normal data, especially for complex models (Brown, 2006; Dolan, 1994; Muthén & Kaplan, 1985).

Several estimation procedures have been developed for the estimation of latent-variable models from non-normally distributed data. Weighted least squares (WLS) estimation is frequently recommended, but several drawbacks to this approach exist, including burdensome sample size requirements and some evidence of bias in Monte Carlo studies (Brown, 2006; Chou & Bentler, 1995). So-called “robust weighted least squares” procedures have been shown to perform well in simulation studies (e.g., Flora & Curran, 2004), and normal theory generalized least squares (GLS) estimation has also shown advantages (Muthén & Kaplan, 1992). Research also suggests that asymptotically distribution-free methods may be appropriate (Browne, 1984).

The usefulness of any of these methods, however, is heavily dependent on the particulars of the model being estimated and on the data being used (Flora & Curran, 2004), decreasing our *a priori* confidence in any particular estimation procedure for the current data.

Lenzenweger and colleagues (2001), in their factor analysis of the 57-item IPO, attempted to solve this problem by using a variety of estimation methods, including maximum likelihood estimation, maximum likelihood estimation with corrected chi-square statistics, and iteratively reweighted generalized least squares estimation. In their analysis, these procedures yielded roughly equivalent results (although, as noted above, the reasons for this convergence are unclear). The present investigation employed a similarly pluralistic approach. Procedures used included maximum likelihood estimation based on a polychoric correlation matrix, estimation using methods for asymptotically distribution-free covariance structures, WLS and GLS estimation, and robust weighted least squares (Flora & Curran, 2004).

Evaluating model fit

One advantage of confirmatory factor analysis over more traditional latent-variable techniques (e.g., exploratory factor analysis or principal components analysis) is that a null hypothesis statistical test can be performed. These tests evaluate the null hypothesis that the expected covariance structure given the postulated model matches the observed covariance structure obtained from the data. The test statistic for this analysis follows a chi-square distribution (Brown, 2006) and can be conceived of as a "goodness-of-fit" test. In this context, the power to detect model misfit is a function of several factors, including model specification, sample size, and model complexity. Because the hypothesized structure of the IPO involves quite a large model, and because the present samples are also large, chi-square goodness-of-fit tests likely have adequate power to reject the null model for reasons entirely separate from the relative fit of its hypothesized factor model (which is the substantive question of interest). Indeed, in their sample of only 249, Lenzenweger and colleagues (2001) obtained highly significant chi-square values for each of their factor models (all *p*-values less than 0.001). This suggests that the complexity of the model alone is enough to cause easily detectable misfit with much smaller sample sizes than the current samples.

In addition, several commentators (Brannick, 1995; Cheung & Rensvold, 2002; Kelloway, 1995) have pointed out that chi-square difference tests are also vulnerable to inflation due to sample size. Therefore, chi-square values and chi-square difference tests were not used as primary tests of the fit of two-and three-factor models in the current study. Rather, a multiple-index strategy was adopted, as recommended by Bentler (1990) and Hu and Bentler (1999), wherein several other fit indices that are relatively insensitive to the size of models and samples were used to evaluate model fit. These statistics included the Comparative Fit Index (CFI), Non-Normed Fit Index (NNFI), and the Standardized Root Mean Square Residual (SRMR), the combination of which is intended to correct for the particular biases of the individual fit indices when using maximum likelihood estimation (Bentler, 1990; Hu & Bentler, 1999). A model was considered adequate if it had a CFI above 0.95, an NNFI above 0.95, and an SRMR below 0.08. These fairly strict standards for the CFI and NNFI were used in the current study because the comparatively lenient, but still commonly used, cutoffs of 0.90 are shown by Hu and Bentler to result in high Type II error rates with large samples. That is, too few improperly specified models are rejected when these values are adopted as cutoffs for large sample sizes.

Measurement equivalence

The current project also seeks to evaluate whether the IPO measures the same constructs, with the same degree of precision, across two different samples. In order to investigate this, a formal test of measurement equivalence across the current samples was conducted. However, the methodological literature contains substantial disagreement over the proper standards for demonstrating measurement equivalence (for a review, see Vandenburg & Lance, 2000). Some authors consider measurement to be equivalent across groups only if all parameters are equal, and others accept invariance of measurement if some subset can be considered equal. Parameters thought by various authors to be important for measurement equivalence include the overall factor structure, factor loadings, item means, item variances, factor means, factor variances, and factor covariances (Vandenburg & Lance, 2000). The importance of equivalence in any of these parameters depends on the theoretical interests of the researchers.

For the purposes of the current study, we tested for the so-called "metric invariance" condition (Horn & McArdle, 1992), under which both the overall model structure and individual factor loadings are equivalent. This standard was used because the primary question of interest is whether the items on the IPO measure the same latent constructs with similar precision across samples. Thus, other properties of the observed and latent variables across samples (such as the means of the items or factors) are largely irrelevant. The test was conducted by first fitting a factor model to both groups simultaneously and then constraining the factor loadings so that they are equal across groups, as recommended by Vandenburg and Lance (2000). A chi-square difference test is the usual method for comparing these nested models; however, as with the comparisons between two-factor and three-factor baseline models, mathematical (Brannick, 1995; Kelloway, 1995) and empirical (Meade & Lautenschlager, 2004) work suggests that these difference values are inflated at large absolute values of chi-square. Therefore, for the test of measurement equivalence, restricted and unrestricted models were compared using criteria recommended by Cheung and Rensvold (2002). These authors recommend examining the difference in the CFI, McDonald's Noncentrality Index (NCI), and gamma-hat fit index between restricted and unrestricted models and comparing these differences to cutoff values derived on the basis of a simulation study. Cheung and Rensvold's cutoff values for testing a metric invariance hypothesis, based on an alpha level of .01, are 0.0085 for the CFI, 0.016 for the NCI, and 0.0008 for gamma hat.

The two samples in the present study differ in several potentially important ways, making the current analysis a relatively strong test of measurement equivalence. For example, one is predominantly Caucasian, the other much more ethnically diverse; one is mostly urban, the other suburban or rural; participants in one sample completed the IPO and other measures using the traditional pencil-and-paper method, the other using survey software on the Internet. To the extent that the IPO shows similarities in latent structure between these two groups, then, its psychometric properties may be said to generalize to fairly diverse samples.

However, given the complexity of the hypothesized factor model (46 indicators and 3 intercorrelated latent variables), any hypothesis of equality is likely to be rejected due to a statistically reliable difference between the two samples in one or more model parameters. In such a circumstance, it is especially important to investigate whether any divergent parameter values seem to show any substantive pattern, or whether the difference in parameter values would more properly be characterized as random. Therefore, an exploratory analysis of "partial measurement equivalence," as suggested by Byrne, Shavelson, & Muthén (1989), was designated as the *post hoc* strategy in this situation. This exploration is essentially analogous to a sensitivity analysis, wherein the overall fit will be examined for its sensitivity to differences in factor loadings between samples.

Criterion relations

In order to investigate the criterion relations of the IPO and its subscales with theoretically related constructs, hierarchical regression analyses were conducted between the IPO and the aforementioned measures of self-concept, defenses, affect, and reckless and self-injurious behavior. The strong relationship between identity diffusion and primitive defenses subscales of the IPO in previous research (Lenzenweger et al., 2001; Normandin et al., 2002), as well as the strong theoretical relationship between the two constructs, argues against an *a priori* causal ordering of these variables in regression analysis. Therefore, each hierarchical regression analysis was conducted twice: first, the identity diffusion subscale was entered as the first predictor and primitive defenses as the second; then, the order of entry was reversed.

Hypotheses

Factor structure of the IPO

Based on previous research on different versions of the IPO (Lenzenweger et al., 2001; Normandin et al., 2002; Oldham et al., 1985), which all showed high intercorrelations between identity diffusion and primitive defenses, we hypothesized that a two-factor model, with identity diffusion and primitive defenses items loading on a common latent factor, would provide a better fit than a model in which identity diffusion, primitive defenses, and reality testing load on three separate factors.

Measurement equivalence

We hypothesized that the IPO would be found to measure similar constructs, with similar patterns of factor loadings, across the two samples in the current study. Statistically, this analysis examined the differences in fit indices between a model with factor loadings constrained to be equal across samples and a model where factor loadings are unrestricted in each sample (but have the same pattern of fixed and free loadings). The null hypothesis in this case was that there would be no difference in fit between the restricted and unrestricted models. As noted above, however, we considered this null hypothesis extremely likely to be rejected, if only for reasons of extreme power and model complexity. Therefore, we hypothesized that at least *partial* measurement equivalence would be found between the current samples, with no discernable pattern in the parameter values responsible for misfit. Because there is no null hypothesis statistical test associated with this hypothesis, any analysis of partial measurement equivalence would be considered strictly exploratory (Byrne, Shavelson, & Muthén, 1989).

Criterion relations

We hypothesized that identity diffusion and primitive defenses will indeed show discriminant criterion relations. Specifically, we hypothesized that identity diffusion would contribute uniquely to the prediction of variance in measures of normative self-concept structure (SCCS, SCM, SSS, LPI, DSI, and BPI identity diffusion), whereas primitive defenses would not. Likewise, we hypothesized that primitive defenses would uniquely predict variance in measures of defense mechanisms (DSQ, BPI primitive defenses, and the Splitting Scale), whereas identity diffusion would not. Finally, we hypothesized that primitive defenses would relate more strongly to measures of affect and impulsive behavior (PANAS, ALS, AIM, DSHI, and RBQ) than would identity diffusion.

Chapter 3. RESULTS

Table 3 presents the means for the Penn State and Hunter College samples for the three primary clinical scales of the IPO.

Table 3

Descriptive Statistics for IPO Clinical Scales

	Penn State			Hunter College		
Scale	*N*	*M*	*SD*	*N*	*M*	*SD*
Identity Diffusion	1153	45.06	10.98	1380	47.98	11.64
Primitive Defenses	1172	31.18	7.77	1380	33.99	8.28
Reality Testing	1190	22.51	7.43	1378	26.08	8.27

Note. Identity diffusion scale score is based on 19 items, primitive defenses on 14 items, and reality testing on 13 items.

Missing data

Because structural equation modeling requires data with no missing values (Brown, 2006), cases with missing data were filled in using LISREL's multiple imputation algorithm. In this technique, the response patterns in the overall data set were used to impute the most likely value for each missing response using a Markov Chain Monte Carlo method. In order to preserve the distributional structure of the data, these imputed values were then rounded to the nearest response category used in the IPO (1, 2, 3, 4, or 5), as recommended by Schafer (1997). 89 data points were imputed for the Hunter College sample, and 65 data points were imputed for the Penn State sample, making up approximately 0.1% of the total data for each group.

Model fit

As noted above, it was difficult to tell *a priori* which estimation technique was most suitable for the current analyses because different factors (e.g., model complexity, sample size, degree of misspecification, and deviations from normality) lead to non-convergence and biased estimates for each technique. Therefore, multiple estimation techniques were used to model the data in order to arrive at a factor solution robust to deviations from normality. Maximum likelihood (ML), weighted least squares (WLS), unweighted least squares (ULS), generalized least squares (GLS), and diagonally weighted least squares (DWLS) estimation techniques were used in LISREL, and robust weighted least squares (WLSMV) was conducted using Mplus.

With the exception of GLS, all techniques were able to converge at an interpretable factor solution. As detailed above, the failure of GLS estimation to produce a convergent solution likely reflects an intolerance of this method for one or more facets of the model or data, both of which have properties that challenge the execution of confirmatory factor analysis. Nevertheless, the fact that all but one method of estimation arrived at a satisfactory factor solution suggests that the CFA results are robust and do not depend on the estimation method used. Tables 4 and 5 summarize the results of the confirmatory factor analyses.

As expected, the chi-square values for all models were highly significant, indicating deviations from ideal fit. A chi-square difference test showed that a three-factor model resulted in significantly less misfit than a two-factor solution ($\Delta\chi^2(3) = 1184$, $p < .001$) for the Penn State

Table 4

Goodness-of-Fit Indices for 2- and 3-factor Models (Hunter College Sample)

Model	χ^2	Satorra-Bentler χ^2	df	CFI	NNFI	SRMR
ML						
2-factor	9171	5252	989	0.96	0.96	0.096
3-factor	8178	4605	986	0.97	0.97	0.068
WLS						
2-factor	26227	-	989	0.66	0.64	0.18
3-factor	10547	-	986	0.87	0.86	0.39
ULS						
2-factor	9369	5375	989	0.96	0.96	0.088
3-factor	8432	4762	986	0.96	0.97	0.058
DWLS						
2-factor	9616	5514	989	0.96	0.96	0.087
3-factor	8627	4867	986	0.97	0.96	0.057
WLSMV						
2-factor	2814	-	349	0.77	0.94	-
3-factor	2830	-	340	0.77	0.94	-

Note. ML = maximum likelihood; WLS = weighted least squares; ULS = unweighted least squares; DWLS = diagonally weighted least squares; WLSMV = “robust” weighted least squares; CFI = comparative fit index; NNFI = non-normed fit index; SRMR = standardized root mean square residual. Generalized least squares estimation did not converge at a factor solution. All estimates derived from LISREL except WLSMV (Mplus). WLS and WLSMV do not return Satorra-Bentler-scaled chi-square values, and WLSMV does not return SRMR fit indices. Cutoffs for adequate model fit: CFI > 0.95; NNFI > 0.95; SRMR < 0.08.

Table 5

Goodness-of-Fit Indices for 2- and 3-factor Models (Penn State Sample)

Model	χ^2	Satorra-Bentler χ^2	df	CFI	NNFI	SRMR
ML						
2-factor	11087	5453	989	0.97	0.97	0.11
3-factor	9903	4467	986	0.97	0.97	0.077
WLS						
2-factor	42030	-	989	0.75	0.74	0.27
3-factor	13759	-	986	0.92	0.92	0.44
ULS						
2-factor	11473	5642	989	0.97	0.96	0.11
3-factor	10369	4985	986	0.97	0.97	0.068
DWLS						
2-factor	11953	5943	989	0.96	0.96	0.10
3-factor	10607	5144	986	0.97	0.97	0.064
WLSMV						
2-factor	2662	-	279	0.75	0.94	-
3-factor	2659	-	280	0.75	0.94	-

Note. ML = maximum likelihood; WLS = weighted least squares; ULS = unweighted least squares; DWLS = diagonally weighted least squares; WLSMV = “robust” weighted least squares; CFI = comparative fit index; NNFI = non-normed fit index; SRMR = standardized root mean square residual. Generalized least squares estimation did not converge at a factor solution. All estimates derived from LISREL except WLSMV (Mplus). WLS and WLSMV do not return Satorra-Bentler-scaled chi-square values, and WLSMV does not return SRMR fit indices. Cutoffs for adequate model fit: CFI > 0.95; NNFI > 0.95; SRMR < 0.08.

sample, as did a chi-square difference test for the Hunter College sample ($\Delta\chi^2(3) = 993$, $p < .001$). As noted above, however, because large models and large samples provide sufficient power to detect small deviations from fit in otherwise satisfactory models (potentially leading to the rejection of the measurement model for scientifically trivial reasons), fit indices that are relatively independent of these factors (that is, CFI, NNFI, and SRMR) were selected *a priori* as the primary statistical basis for evaluating fit. In contrast to the hypotheses of the current study, an examination of the relative fit of three-factor and two-factor models generally supported a three-factor solution for both Penn State and Hunter College samples, with separate factors for reality testing, identity diffusion, and primitive defenses subscales. CFI and NNFI values were higher (indicating better fit) for the three-factor model than for the two-factor model regardless of estimation technique. SRMR values were lower (indicating better fit) for the three-factor model using every technique except weighted least squares regression. There were essentially no differences between models in fit indices using robust weighted least squares. Therefore, although the results were not entirely consistent, the prevailing statistical evidence suggests that a three-factor solution is preferable to a two-factor solution for the current data.

This conclusion is further supported by an examination of the absolute fit of the models.[1] The three-factor model had adequate fit indices with respect to the cutoffs suggested by Hu and Bentler (1999) for maximum likelihood estimation for both the Penn State (CFI = .97, NNFI = .97, SRMR = .077) and Hunter College (CFI = .97, NNFI = .97, SRMR = .068) samples, whereas the two-factor model ($CFI_{PSU} = .97$, $NNFI_{PSU} = .97$, $SRMR_{PSU} = .11$; $CFI_{Hunter} = .96$,

[1] Note that only the SRMR is an index of “absolute fit,” whereas the CFI and NNFI depend on a comparison with a baseline model. Here, however, we use the term “absolute fit” to refer to the models’ fit with respect to our numerical cutoffs for all three indices (as opposed to the “relative fit” of two- and three-factor models as compared to one another).

$NNFI_{Hunter}$ = .96, $SRMR_{Hunter}$ = .096) did not. Specifically, although both models had adequate CFI and NNFI values (above 0.95) in each sample, only the three-factor model had SRMR values below the recommended cutoff of 0.08.[2] On the basis of these results, a three-factor solution was accepted for each sample. The correlations between three latent factors were generally high: identity diffusion and primitive defenses = .98, primitive defenses and reality testing = .70, identity diffusion and reality testing = .66. These values are consistent with the factor intercorrelations found by Lenzenweger et al. (2001) in their three-factor model (.97, .71, and .67).

Measurement equivalence

Because the factor structure of the IPO was found to be consistent across groups, more specific hypotheses concerning this measurement model could be tested. Specifically, the "metric invariance" of the three-factor model was tested across the Penn State and Hunter College samples. To do this, a multigroup CFA was performed using the three-factor structure identified in the analysis above. Two models were tested: one in which the factor loadings were held equivalent across groups, and one in which these factor loadings were allowed to vary freely within the specified structure. As above, each IPO item was allowed to load only on one latent factor (that is, the other potential loadings were held at zero).

Based on Cheung and Rensvold's (2002) criteria, a hypothesis of wholesale measurement equivalence between groups was rejected. Although the differences in CFI (0.0061) and NCI (0.00057) indices were well below cutoffs, the difference in gamma hat between models was too

[2] Hu and Bentler's (1999) cutoffs pertain to fit indices derived using maximum likelihood estimation. They are used here because, to our knowledge, there are no widely accepted, empirically derived cutoffs for use with other estimation methods.

high (0.0024) for a nominal alpha level of 0.01. Therefore, we proceeded to test for partial measurement equivalence using the procedure suggested by Byrne, Shavelson, & Muthén (1989). Using LISREL, we began by removing the equality-between-groups constraint on the factor loading with the highest modification index. The resultant model was evaluated and compared to the completely unrestricted model. This model did not reduce the difference in gamma hat sufficiently, and so a second factor loading was allowed to vary between groups according to the modification indices. This procedure was repeated until the numerical cutoff of 0.0008 was reached for the gamma hat index. At the end of this sequence, nine factor loadings had been allowed to vary between samples. Table 6 provides a list of the items whose factor loadings differed, presented in the order in which they were allowed to vary. Table 6 also presents the differences in gamma hat between the model in which the factor loadings for that item and all those above it were allowed to vary between groups and the model in which all factors loadings were held equivalent between groups.

Four items whose factor loadings were eventually allowed to vary between groups were from the reality testing subscale, three were from the primitive defenses subscale, and two were from the identity diffusion subscale. There was no discernable pattern to these items in the scales they belonged to or in the degree to which they contributed to measurement variance. Therefore, a tentative conclusion of partial measurement equivalence was accepted for these data. It appears that the loadings of IPO items onto their theoretically specified factors (identity diffusion, primitive defenses, and reality testing) are largely the same between these groups.

Criterion relations

Our final hypotheses concerned the external validity of IPO subscales. In particular, we sought to test whether identity diffusion and primitive defenses scales would show theoretically

Table 6

Items Allowed to Vary Between Samples with Resultant Gamma-hat Difference from Model with Invariant Factor Loadings

Item	Factor	Item Content	Δγ-hat
95	RT	I have seen things which do not exist in reality.	.00212
17	RT	When I'm nervous or confused, it seems like things in the outside world don't make sense either.	.00189
78	ID	It is hard for me to be sure about what others think of me, even people who have known me very well.	.00159
102	PD	I find myself doing things which feel okay while I am doing them but which I later find hard to believe I did.	.00141
166	PD	I feel I don't get what I want.	.00126
51	PD	I find myself doing things which at other times I think are not too wise, like having promiscuous sex, lying, drinking, having temper tantrums, or breaking the law in minor ways.	.00110
74	RT	I understand and know things that nobody else is able to understand or know.	.00099
47	RT	I have seen or heard things when there is no apparent reason for it.	.00089
83	ID	In the course of an intimate relationship, I'm afraid of losing a sense of myself.	.00078

Note. Items are listed in the order in which their factor loadings were freed. The first item allowed to vary is at top. RT = reality testing; PD = primitive defenses; ID = identity diffusion.

appropriate patterns in their relationships with other measures of self-concept structure and defenses. Some of the criterion scales have been used fairly infrequently, and there is only limited data regarding their psychometric properties. However, an examination of these scales' internal consistency coefficients (Cronbach's α) in the current study revealed that all scales had adequate homogeneity ($\alpha > 0.70$). Table 7 presents these values.

Because identity diffusion and primitive defenses subscales have generally been found to be highly intercorrelated (Lenzenweger et al., 2001; Normandin et al., 2002), there was little *a priori* reason to suggest one ordering of predictors over another in the regression hierarchy, despite the hypothesized relationships. Thus, two forced-entry regression analyses were conducted for each criterion variable, one in which identity diffusion was entered as the first predictor in the model and primitive defenses entered second, and another analysis in which the order of entry was reversed. Tables 8 and 9 present the results of these analyses.

Identity measures

Consistent with hypotheses, the identity diffusion subscale of the IPO was a better predictor of the measures of identity structure than the primitive defenses subscale. Identity diffusion predicted variance in each criterion variable above and beyond primitive defenses, whereas primitive defenses contributed uniquely to the prediction of only two scales: the identity diffusion subscale on the Borderline Personality Inventory and Ziller's Self-Complexity Measure (SCM). IPO identity diffusion showed no zero-order correlation with the SCM, suggesting that this scale related differently to the IPO subscales than the other measures of self-concept structure. Except for the SCM, the identity diffusion and reality testing subscales together predicted between 30% and 50% of the variance in the identity scales. The identity diffusion subscale accounted for most of this common variance.

Table 7

Internal Consistency Coefficients (Cronbach's α*)*

Scale	Cronbach's α
IPO identity diffusion	
Penn State	0.876
Hunter College	0.856
IPO primitive defenses	
Penn State	0.819
Hunter College	0.786
IPO reality testing	
Penn State	0.866
Hunter College	0.844
Self-Concept Clarity Scale (SCCS)	0.859
Borderline Personality Inventory (BPI) identity diffusion	0.804
Stability of Self Scale (SSS)	0.762
Life Problems Inventory (LPI) confusion about self	0.944
Differentiation of Self Inventory (DSI)	0.957
Splitting Scale	0.704
Borderline Personality Inventory (BPI) primitive defenses	0.755
Defense Styles Questionnaire (DSQ) immature defenses	0.835
Positive and Negative Affect Schedule (PANAS) positive	0.863
Positive and Negative Affect Schedule (PANAS) negative	0.833
Affect Lability Scale (ALS)	0.973
Affect Intensity Measure (AIM)	0.928
Reckless Behavior Questionnaire (RBQ)	0.875

Table 8

Standardized Regression Weights (β) and R^2 *Change Values for Identity Diffusion (ID; Entered in Block 1) and Primitive Defenses (PD; Added in Block 2) in Hierarchical Regression Analyses*

		Block 1	Block 2			
Scale	*N*	ID	ID	PD	ΔR^2	R^2_{total}
Identity						
SCCS	1200	-.683***	-.718***	.043	.001	.467***
BPI_{ID}	1202	.588***	.417***	.208***	.014***	.360***
SCM	721	.014	-.203***	.260***	.020***	.020***
SSS	744	-.559***	-.595***	.059	.001	.314***
LPI_{CAS}	268	.638***	.484***	.178	.008	.415***
DSI	263	-.574***	-.592***	.021	.001	.330***
Defense						
Splitting	1253	.589***	.421***	.210***	.016***	.363***
BPI_{PD}	1202	.551***	.303***	.303***	.030***	.334***
DSQ_{PD}	1363	.493***	.175***	.396***	.056***	.299***
Affect						
$PANAS^{+}$	1248	-.179***	-.201***	.027	< .001	.032***
$PANAS^{-}$	1248	.488***	.286***	.252***	.023***	.261***
ALS	1206	.075**	.021	.065	.001	.007*
AIM	1199	.144***	-.143**	.350***	.040***	.061***
Behavior						
RBQ	1201	.137***	-.033	.208***	.014***	.033***
DSHI[a]	1210	2.212***	1.674*	1.43		

Note. *: $p < .05$; **: $p < .01$; ***: $p < .001$. SCCS = Self-Concept Clarity Scale; BPI_{ID} = Borderline Personality Inventory identity diffusion; SCM = Self-Complexity Measure; SSS = Stability of Self Scale; LPI_{CAS} = Life Problems Inventory confusion about self; DSI = Differentiation of Self Inventory; Splitting = Splitting Scale; BPI_{PD} = Borderline Personality Inventory primitive defenses; DSQ_{PD} = Defense Styles Questionnaire immature (primitive) defenses; $PANAS^{+}$ = Positive and Negative Affect Schedule positive items; $PANAS^{-}$ = Positive and Negative Affect Schedule negative items; ALS = Affective Lability Scale; AIM = Affect Intensity Measure; RBQ = Reckless Behavior Questionnaire; DSHI = Deliberate Self-Harm Inventory. All regressions were carried out using scale means.

[a]The regression weights for the DSHI are given in terms of the zero-order and partial odds ratios.

Table 9

Standardized Regression Weights (β) and R^2 *Change Values for Primitive Defenses (PD; Entered in Block 1) and Identity Diffusion (ID; Added in Block 2) in Hierarchical Regression Analyses*

		Block 1	Block 2			
Scale	*N*	PD	PD	ID	ΔR^2	R^2_{total}
Identity						
SCCS	1200	-.546***	.043	-.718***	.169***	.467***
BPI_{ID}	1202	.550***	.208***	.417***	.050***	.360***
SCM	721	.089*	.260***	-.203***	.012***	.020***
SSS	744	-.470***	.059	-.595***	.105***	.314***
LPI_{CAS}	268	.596***	.178	.484***	.060***	.415***
DSI	263	-.490***	.021	-.592***	.090***	.330***
Defense						
Splitting	1253	.547***	.210***	.421***	.063***	.363***
BPI_{PD}	1202	.551***	.303***	.303***	.030***	.334***
DSQ_{PD}	1363	.537***	.396***	.175***	.011***	.299***
Affect						
$PANAS^{+}$	1248	-.134***	.027	-.201***	.014***	.032***
$PANAS^{-}$	1248	.482***	.252***	.286***	.029***	.261***
ALS	1206	.083**	.065	.021	< .001	.007*
AIM	1199	.232***	.350***	-.143**	.007	.061***
Behavior						
RBQ	1201	.180***	.208***	-.033	.001	.033***
DSHI[a]	1210	2.212***	1.43	1.674*		

Note. *: p < .05; **: p < .01; ***: p < .001. SCCS = Self-Concept Clarity Scale; BPI_{ID} = Borderline Personality Inventory identity diffusion; SCM = Self-Complexity Measure; SSS = Stability of Self Scale; LPI_{CAS} = Life Problems Inventory confusion about self; DSI = Differentiation of Self Inventory; Splitting = Splitting Scale; BPI_{PD} = Borderline Personality Inventory primitive defenses; DSQ_{PD} = Defense Styles Questionnaire immature (primitive) defenses; $PANAS^+$ = Positive and Negative Affect Schedule positive items; $PANAS^-$ = Positive and Negative Affect Schedule negative items; ALS = Affective Lability Scale; AIM = Affect Intensity Measure; RBQ = Reckless Behavior Questionnaire; DSHI = Deliberate Self-Harm Inventory. All regressions were carried out using scale means.
[a]The regression weights for the DSHI are given in terms of the zero-order and partial odds ratios.

Defense measures

The relationships between IPO subscales and measures of defenses were mixed. Consistent with hypotheses, the IPO primitive defenses subscale predicted unique variance in all three defense scales. However, the identity diffusion subscale also contributed to the prediction of these scales above and beyond primitive defenses. In particular, identity diffusion (partial β = .421, p < .001) was more strongly related to Gerson's Splitting Scale than primitive defenses (partial β = .210, p < .001) when both predictors were added to the model. An exploratory follow-up analysis showed that this result was not entirely due to Gerson's inclusion of an identity diffusion item in her scale; without that item, the same result held, although the effect lessened (partial β_{ID} = ..389, p < .001; partial β_{PD} = .216, p < .001).

A different pattern held of relationships held between the IPO subscales and the immature defenses composite scale derived from the Defense Styles Questionnaire. Primitive defenses (partial β = .396, p < .001) was a stronger predictor than identity diffusion (partial β = .175, p < .001) in the full model. Finally, the third measure of defense processes, the primitive defenses subscale of the Borderline Personality Inventory, was identically related to the identity diffusion (zero-order $\beta = .551$, $p < .001$; partial $\beta = .303$, $p < .001$) and primitive defenses (zero-order $\beta = .551$, $p < .001$; partial $\beta = .303$, $p < .001$) subscales of the IPO, according to current estimates. Therefore, the overall hypothesis that the IPO primitive defenses subscale would demonstrate incremental validity in the prediction of other measures of defense was not supported and seemed to depend on the particular criterion scale. Together, the identity diffusion and primitive defenses subscales of the IPO accounted for roughly 30% to 40% of the variance in the defense measures.

Affect measures

The relationships between IPO subscales and measures of affect were also mixed. Identity diffusion ($\beta = -.179$, $p < .0010$) and primitive defenses ($\beta = -.134$, $p < .001$) were both negatively related to the positive affect scale of the PANAS, but only identity diffusion (partial $\beta = -.201$, $p < .001$) emerged as a significant predictor when both subscales were entered into the model. In contrast, both subscales contributed uniquely to the prediction of the PANAS negative affect scale (partial $\beta_{ID} = .286$, $p < .001$; partial $\beta_{PD} = .252$, $p < .001$). Both scales showed a significant but trivial relationship with affective lability ($\beta_{ID} = .075$, $p < .01$; $\beta_{PD} = .083$, $p < .01$), but neither scale was significantly related to the ALS when controlling for the other. Finally, both scales contributed uniquely to the prediction of variance in affect intensity as measured by the AIM (partial $\beta_{ID} = -.143$, $p < .01$; partial $\beta_{PD} = .350$, $p < .001$). The overall level of variance in measures of affect accounted for by the IPO subscales was typically very small to small (between 1% and 26%).

Behavior measures

The relationship of measures of reckless and self-harming behavior with IPO subscales also showed a mixed pattern. Consistent with hypotheses, the primitive defenses subscale of the IPO was the stronger predictor of scores on the Reckless Behavior Questionnaire. Whereas both scales predicted reckless behavior independently (β_{ID} = .137, p < .001; β_{PD} = .180, p < .001), only primitive defenses predicted unique variance in reckless behavior when both subscales were entered into the model (partial β_{PD} = .208, p < .001). The overall amount of variance in the RBQ accounted for by the IPO subscales was very small (3%).

Consistent with their non-clinical status, approximately 81% of the Penn State sample reported no history of deliberate self-harm on the DSHI. In order to accommodate the healthy nature of the sample, DSHI responses were coded as 1 (lifetime history present) and 0 (lifetime history absent), and these scores were regressed onto IPO subscales using hierarchical logistic regression (Cohen, Cohen, West, & Aiken, 2003). Results showed that both subscales predicted the lifetime incidence of self-harm independently (OR_{ID} = 2.212, p < .001; OR_{PD} = 2.184, p < .001), but only identity diffusion remained a significant predictor of self-harm in the full model (partial OR = 1.674, p < .05). This finding was the opposite of the hypothesized relationship.

Chapter 4. DISCUSSION

Model Fit

In contrast to Lenzenweger et al. (2001), who adopted a two-factor measurement model for the larger version of the IPO, the current study found that a three-factor solution was optimal for the 46-item version of the scale. There are several possible reasons for this discrepancy. First, the current study used a much larger sample than that used by Lenzenweger and colleagues. Therefore, it could be that the current factor solution is simply a better approximation of an underlying model that is common to both studies.

Second, it could be that the use of different fit criteria resulted in different conclusions. Lenzenweger et al. (2001) used only χ^2 and AIC fit indices to test their factor models, and none of those models showed good fit to the data based on these criteria. They adopted a two-factor solution, despite finding that a three-factor solution fit better based on a χ^2 difference test, for reasons of relative fit and parsimony. In contrast, the current study designated, *a priori*, absolute fit index cutoffs that were used to evaluate model fit; these criteria were a CFI and NNFI greater than 0.95 and SRMR values below 0.08. The three-factor solution was the only model of the two that provided a plausible fit based on these cutoffs.

A third possible reason for the differences between the factor structure obtained for the current data and that of Lenzenweger et al. (2001) is that different versions of the IPO were used in the two studies. Whereas Lenzenweger and colleagues used a version with 57 items, a 46-item version was used in the current study. Although the shorter version was designed to measure the same constructs, it is likely that omitting 11 items altered the factor structure to some extent. Using the current data, it is impossible to evaluate the degree to which differences in factor structure are due to the use of different IPO versions, but future studies may be able to address this discrepancy by attempting to replicate both structures in a single sample.

The present study has a number of limitations that caution against a strong conclusion that a three-factor model characterizes the IPO well in all contexts. First, as shown in Tables 4 and 5, this conclusion depended to some degree on the method of estimation used, with at least one estimation technique (WLS) seeming to favor a two-factor model. Because so many factors affect the parameter estimates given by different estimation methods, it is difficult to tell which method is “best” for any given data set. Therefore, it is possible that the present conclusion is partly an artifact of choosing one method over another.

Second, the factor solution adopted for the present data is somewhat dependent on the criteria used to evaluate model fit. The differences in fit between two- and three-factor models were not large, and an argument could be made that both models fit the data plausibly. Whereas the cutoffs recommended by Hu and Bentler (1999) led to the conclusion that a three-factor model was preferable, less stringent cutoffs or different fit indices may have resulted in a more ambiguous conclusion. Altogether different concerns (such as parsimony) may also have resulted in different decisions regarding model fit.

In addition, the current study only evaluated two competing models. There may be other structures that provide a better fit to the data. For example, a hierarchical model in which identity diffusion and primitive defenses factors are subordinate to a higher-order “BPO” factor would perhaps be theoretically appropriate given Kernberg’s (1975) theory of borderline personality, in which these constructs are tightly linked. This model would also have the potential benefit of accounting for the strong association between the identity diffusion and primitive defenses factors in an explicit and interpretable way. However, such a structure would add even more complexity to the analysis and would be difficult to evaluate without a large sample.

Finally, as in any factor analysis, the current results depend to some degree on the sample used. One important consideration is that the two samples used were non-clinical and composed of college undergraduates. Although this is a common characteristic of factor analytic studies (and characterizes every factor analysis of the IPO to the present date), it is certainly possible that the IPO has different psychometric properties in other populations. Investigations of this possibility in diverse samples, including clinical samples and samples not composed of college undergraduates, will be important in evaluating the generalizability of this structure for the IPO.

Measurement Equivalence

According to the analysis of measurement equivalence, the strict hypothesis that the IPO measures the same constructs with equal precision across the two samples was rejected. A further analysis demonstrated that the relationship between 9 IPO items and latent identity, defense, and reality testing factors varied between groups, whereas the other 37 items were equivalent indicators of these factors across groups. This situation might be termed “partial measurement equivalence” (Byrne, Shavelson, & Muthén, 1989). It does not appear that the items with varying factor loadings come from one subscale over the others or share any similarity in terms of content.

However, it must be emphasized that the analysis of partial measurement equivalence should be considered exploratory. Replication of this analysis will be needed to show whether these factor loadings show robust differences between groups or whether they vary by chance across the current samples. If these items show replicable differences, the IPO might be improved by altering the content of those items to eliminate any bias, and the question of why those items tended to be better measures of the IPO's theoretical constructs in certain groups would be important to investigate.

It should be noted that any conclusion of measurement equivalence between groups depends on the groups being compared and is not an indication of the universal measurement equivalence of the questionnaire itself. Further research will be needed in order to determine whether this factor structure of the IPO generalizes to other groups (for example, clinic outpatients or older adults). It should also be noted that the test of measurement equivalence in the current study only concerned the factor loadings, not other parameters that might be of interest, such as factor means, factor variances, or item means.

Criterion Relations

An examination of the relationships of IPO subscales with external measures was important for several reasons. First, whereas Lenzenweger et al. (2001) provided a test of the external validity of the reality testing subscale in their larger version of the IPO, they did not evaluate the relationship between the identity diffusion and primitive defenses subscales and other measures of self-concept structure and defenses. Second, because the factor analysis showed that a three-factor model was a good fit to the data but did not unambiguously rule out other models, it is important to examine whether the identity diffusion and primitive defenses scales have incremental validity in relation to these constructs. To the extent that other measures do not differ in their relationship with IPO identity diffusion and primitive defenses, there may be little practical value in separating the two subscales. In addition, because the large sample and large model rendered a strict null hypothesis statistical test for the factor analysis hard to interpret, an examination of the external validity of different parts of the model can bolster confidence in the fit indices that were obtained (Barrett, 2007).

IPO identity diffusion showed strong convergent validity with other measures of self-concept structure. Of the six self-concept measures included in the current study, the identity diffusion subscale predicted variance in all of them above and beyond the primitive defenses subscale. Moreover, the identity diffusion subscale entirely accounted for the relationships between these self-concept measures and the primitive defenses subscale except in two cases: the identity diffusion subscale of the Borderline Personality Inventory (BPI) and Ziller's Self-Concept Measure (SCM). In the case of the BPI identity diffusion subscale, the continued variance shared with the IPO primitive defenses scale after accounting for the variance shared with IPO identity diffusion may be a function of those 3 items on the BPI identity diffusion scale (out of 10 total) found to load strongly on a "dissociative/psychotic symptoms" factor by Chabrol and colleagues (2004, p. 61). These items could plausibly relate to a defensive process, despite being designed by Leichsenring (1999) to relate only to identity diffusion.

The fact that the IPO identity diffusion subscale did not show a zero-order relationship with self-complexity, despite adequate power to detect even small effects in the current sample, is noteworthy and demands explanation. One possibility is that the lack of common assessment method between the two measures attenuated their relationship. Because self-complexity was measured in the present study using a free checklist of self-attributes (the SCM), it may have been expressed differently than it would have been using a questionnaire-type measure with Likert-scale responses. Another possibility is that self-complexity has a more complicated relationship with identity diffusion than can be summed up in a single linear regression coefficient. Theoretically, identity diffusion may not relate not to complexity *per se* but to the degree to which aspects of the self-concept are integrated and coherent. An as-yet-unidentified moderator variable (or set of variables) could alter its relationship with complexity under different conditions. Thus, identity diffusion could manifest in a chaotic and fragmented self-concept with many self-aspects, or conversely, in an impoverished and empty self-concept, where none of the proferred self-aspects seems to be true of the person. This confusion between complexity and coherence, which has been noted by self-concept researchers as a problem in both theory and in assessment (for a review, see Rafaeli-Mor & Steinberg, 2002), may contribute to the somewhat surprising result in the present study that identity diffusion and self-complexity were unrelated.

In sum, however, it appears that the identity diffusion subscale shows strong relationships with other measures of self-concept structure. This finding argues against the conjecture of Lenzenweger and colleagues (2001), based on null correlations between identity diffusion and measures of self-monitoring and self-consciousness, that IPO identity diffusion is an indicator of pathological identity as opposed to normative, non-pathological self-concept. Rather, it appears that the identity diffusion scale is specifically a measure of *self-concept structure* (both healthy and pathological) and is unrelated to measures of *self-reflection*, such as self-monitoring and self-consciousness (Lenzenweger et al.).

However, identity diffusion also showed incremental validity above primitive defenses in the prediction of external measures of defense processes. Identity diffusion was a stronger predictor of scores on Gerson's (1984) Splitting Scale than primitive defenses when both subscales were entered into the hierarchical regression analysis. Some of this may be due to the fact that Gerson included one item in the scale that was specifically designed to capture identity diffusion, due to its close relationship to splitting. However, a *post hoc* analysis revealed that this item was not wholly responsible for this pattern. This result suggests that the IPO identity diffusion subscale does not show discriminant validity with respect to splitting but contains some overlap with this construct that is still unexplained.

Moreover, identity diffusion predicted the primitive defenses subscale of the BPI (Leichsenring, 1999) with more or less identical strength as the primitive defenses subscale of the IPO. Again, it is possible that the BPI primitive defenses scale, like the BPI identity diffusion scale, contains some content overlap with identity diffusion. Supporting this notion, Chabrol et al. (2004, p. 61) found that 4 out of the 8 BPI primitive defenses items loaded strongly on a factor they labeled "affectivity/identity disturbance." These items may contribute to the common variance between this subscale and the identity diffusion subscale of the IPO.

As noted above, the primitive defenses subscale of the IPO showed inconsistent evidence of convergent validity with other measures of defense. It had incremental value in predicting all three criterion scales of defense, but in only one case was this relationship stronger than the relationship of the criterion scale to IPO identity diffusion. In some respects, this result is unsurprising, given the generally low correlation between different measures of defense processes (Davidson & MacGregor, 1998). This may be due to the fact that it is, by definition, difficult to directly assess psychological processes that are outside of awareness using self-report questionnaires. A typical strategy in constructing measures of defense mechanisms is to ask the respondent about behaviors and attitudes that might result from their operation. As Davidson and MacGregor (1998) have pointed out, though, the meaning of these behaviors and attitudes is quite idiographic, and the assessment of them may not provide a generally valid indicator of defenses. The three defense questionnaires used as criterion measures in the current study may suffer from this limitation.

Thus, the three measures of defenses used in the current study may not have sufficient validity themselves to be used effectively to validate the primitive defenses subscale of the IPO. Other methods of assessment, such as observer-rated measures of defenses (Perry & Ianni, 1998), assessments of the discrepancy between self-rated and other-rated attributes (Davidson, 1993), or assessments of discrepancy between self-ratings and physiological data (Shedler, Mayman, & Manis, 1993) may be more consistent with theories of defense processes than many self-report measures and may thus be better criteria for validating any new assessment of defenses. In all, the use of the primitive defenses subscale of the IPO to assess the defensive styles of borderline personality remains unjustified based on the present investigation.

The IPO subscales related generally weakly and inconsistently to measures of affect. The identity diffusion subscale, but not primitive defenses, captured unique variance in positive affectivity as measured by the PANAS; the combined variance accounted for by both subscales was small. Both subscales showed predicted incremental variance on the negative subscale of the PANAS, and the association was small to moderate. This discrepancy between the IPO's relationship to positive and negative affect is consistent with substantial empirical work showing that these dimensions of affect are relatively independent (e.g., Tellegen, Watson, & Clark, 1999). However, the small relationship between the IPO subscales and positive affect is inconsistent with the results of Lenzenweger and colleagues (2001), who found a much stronger relationship between all subscales and positive affect on the PANAS.

The identity diffusion and primitive defenses scales related only weakly to affective lability, and neither subscale predicted any incremental variance over the other. Both scales predicted unique variance in affect intensity, however, although the overall proportion of variance accounted for was small. On the one hand, the fact that these relationships were so weak compared to the IPO's relationships with external identity and defense measures is evidence that the IPO does not measure aspects of borderline phenomenology that it was not designed to measure. On the other hand, this result suggests that the IPO is not a good measure of the full breadth of borderline personality, at least in a non-clinical sample.

Finally, the identity diffusion and primitive defenses subscales showed different relationships with two measures of dangerous behavior. While both subscales showed zero-order correlations with the RBQ, only primitive defenses accounted for unique variance in reckless behavior. In contrast, only identity diffusion predicted the lifetime incidence of deliberate self-harm when both subscales were used to predict it. These results suggest that reckless behavior may have different correlates in college students than self-injurious behavior. However, it should be noted that the effect sizes here were relatively small. The IPO subscales predicted only about 3% of the variance in reckless behavior. Similarly, a one-point increase on either identity diffusion or primitive defenses (out of a mean score scaled to the 5-point Likert scale) roughly doubled the odds of lifetime self-harm, which could be considered a small increase considering the overall base rate of such behavior. Thus, while identity diffusion and primitive defenses predict distinct types of risky behavior in a college sample, they may not account for as much of this behavior as other variables, such as childhood maltreatment (Gratz, 2006) or sensation seeking (Wagner, 2001).

Overall Conclusions and Directions for Future Research

Together, results from the factor analysis of the IPO and the patterns of criterion-related validity of the IPO identity diffusion and primitive defenses subscales can be taken as a partial validation of the theoretically-derived substructure of the measure. The three-factor model appears empirically plausible based on CFA results. Moreover, the IPO factor structure seems robust to groups with different demographic characteristics and different methods of administration. It also appears that a three-factor structure results in a valid subscale measuring self-concept structure, which is a core feature of borderline personality. However, the current study did not support the use of the primitive defenses subscale as a measure of the characteristic defensive style seen in borderline personality.

Future research should evaluate the plausibility of other models that might also be theoretically consistent with Kernberg's (1975) theory of personality organization, such as a hierarchical model in which identity diffusion and primitive defenses are represented by distinct subfactors of a general borderline personality factor. Valid measurement models could also be used to construct structural equation models to evaluate causal relationships among personality organization and other indicators of dysfunction. The psychometric properties of the IPO also should be studied in clinical samples in order to see whether the measure is able to assess a more severe range of psychopathology. Finally, it is also clear that more research is needed to validate a borderline-specific measure of defenses that can be used in large-scale research.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4^{th} ed.). Washington, DC: Author.

Andrews, G., Singh, M., & Bond, M. (1993). The Defense Style Questionnaire. *Journal of Nervous and Mental Disease, 181*, 246-256.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*, 815-824.

Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of Borderline Personality Disorder: A randomized controlled trial. *American Journal of Psychiatry, 156*, 1563-1569.

Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., et al. (2001). Treatment utilization by patients with personality disorders. *American Journal of Psychiatry, 158*, 295-302.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Blashfield, R. K., & Livesley, W. J. (1991). Metaphorical analysis of psychiatric classification as a psychological test. *Journal of Abnormal Psychology, 100,* 262-270.

Blatt, S. J., & Levy, K. N. (1998). A psychodynamic approach to the diagnosis of psychopathology. In J. W. Barron (Ed.), *Making diagnosis meaningful: Enhancing evaluation and treatment of psychological disorders* (pp. 73-109). Washington, DC: American Psychological Association.

Bond, M., Paris, J., & Zweig-Frank, H. (1994). Defense styles and borderline personality disorder. *Journal of Personality Disorders, 8*, 28-31.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior, 16*, 201-213.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83.

Bunce, S. C., & Coccaro, E. (1999). Factors differentiating personality-disordered individuals with and without a history of unipolar mood disorder. *Depression and Anxiety, 10*, 147-157.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.

Campbell, J. D., Assanand, S., & Di Paula, A. (2003). The structure of the self-concept and its relation to psychological adjustment. *Journal of Personality, 71*, 115-140.

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*, 141-156.

Carson, R. C. (1991). Dilemmas in the pathway of the *DSM-IV*. *Journal of Abnormal Psychology, 100*, 302-307.

Chabrol, H., Montovany, A., Ducongé, E., Kallmeyer, A., Mullet, E., & Leichsenring, F. (2004). Factor structure of the Borderline Personality Inventory in adolescents. *European Journal of Psychological Assessment, 20*, 59-65.

Chambless, D. L., Renneberg, B., Goldstein, A., & Gracely, E. J. (1992). MCMI-diagnosed personality disorders among agoraphobic outpatients: Prevalence and relationship to severity and treatment outcome. *Journal of Anxiety Disorders, 6*, 193-211.

Cheng, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37-75). Thousand Oaks, CA: Sage.

Clark, L. A. (1993). *Manual for the Schedule for Nonadaptive and Adaptive Personality (SNAP).* Minneapolis: University of Minnesota Press.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.

Clarkin, J. F., Caligor, E., Stern, B., & Kernberg, O. F. (2002). *Structured interview of personality organization (STIPO)*. Unpublished manuscript. New York: Institute for Personality Disorders.

Clarkin, J. F., Foelsch, P. A., & Kernberg, O. F. (2001). *The Inventory of Personality Organization*. White Plains, NY: The New York Hospital-Cornell Medical Center.

Clarkin, J. F., Lenzenweger, M. F., Yeomans, F., Levy, K. N., & Kernberg, O. F. (2007). An object relations model of borderline pathology. *Journal of Personality Disorders, 21*, 474-499.

Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating three treatments for borderline personality disorder: A multiwave study. *American Journal of Psychiatry, 164*, 922-928.

Clarkin, J. F., Widiger, T. A., Frances, A., Hurt, S. W., & Gilmore, M. (1983). Prototypic typology and the borderline personality disorder. *Journal of Abnormal Psychology, 92*, 263-275.

Cloitre, M., & Koenen, K. C. (2001). The impact of borderline personality disorder on process group outcome among women with posttraumatic stress disorder related to childhood abuse. *International Journal of Group Psychotherapy, 51*, 379-398.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cooper, P. J., Coker, S., & Fleming, C. (1996). An evaluation of the efficacy of supervised cognitive behavioral self-help for bulimia nervosa. *Journal of Psychosomatic Research, 40*, 281-287.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245-265.

Davidson, K. W. (1993). Suppression and repression to discrepant self-other ratings: Relations with thought control and cardiovascular reactivity. *Journal of Personality, 61*, 669-691.

Davidson, K., & MacGregor, M. W. (1998). A critical appraisal of self-report defense mechanism measures. *Journal of Personality, 66*, 965-992.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*, 309-326.

Fenigstein, A., Sheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522-527.

Flett, G. L., & Hewitt, P. L. (1995). Criterion validity and psychometric properties of the affect intensity measure in a psychiatric sample. *Personality and Individual Differences, 19*, 585-591.

Fliege, H., Kocalevent, R.-D., Walter, O. B., Beck, S., Gratz, K. L., Gutierrez, P. M., et al. (2006). Three assessment tools for deliberate self-harm and suicide behavior: Evaluation and psychopathological correlates. *Journal of Psychosomatic Research, 61*, 113-121.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466-491.

Gangestad, S., & Snyder, M. (1985). “To carve nature at its joints”: On the existence of discrete classes in personality. *Psychological Review, 92*, 317-349.

Gerson, M.-J. (1984). Splitting: The development of a measure. *Journal of Clinical Psychology, 40*, 157-162.

Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., et al. (2006). Outpatient psychotherapy for borderline personality disorder: Randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry, 63*, 649-658.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43-56.

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., et al. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 69,* 533-545.

Gratz, K. L. (2001). Measurement of deliberate self-harm: Preliminary data on the Deliberate Self-Harm Inventory. *Journal of Psychopathology and Behavioral Assessment, 23*, 253-263.

Gratz, K. L. (2006). Risk factors for deliberate self-harm among female college students: The role and interaction of childhood maltreatment, emotional inexpressivity, and affect intensity/reactivity. *American Journal of Orthopsychiatry, 76*, 238-250.

Harvey, P. D., Greenberg, B. R., & Serper, M. R. (1989). The Affective Lability Scales: Development, reliability, and validity. *Journal of Clinical Psychology, 45*, 786-793.

Hébert, É., Diguer, L., Descôteaux, J., Daoust, J.-P., Rousseau, J.-P., Normandin, L., et al. (2003). The Personality Organization Diagnostic Form (PODF): A preliminary report on its validity and interrater reliability. *Psychotherapy Research, 13,* 243-254.

Henry, C., Mitropoulou, V., New, A. S., Koenigsberg, H. W., Silverman, J., & Siever, L. J. (2001). Affective instability and impulsivity in borderline personality and bipolar II disorders: Similarities and differences. *Journal of Psychiatric Research*, *35,* 307-312.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117-144.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Huebner, E. S., & Dew, T. (1995). Preliminary validation of the Positive and Negative Affect Schedule with adolescents. *Journal of Psychoeducational Assessment, 13,* 286-293.

Jackson, D. N. (1970). *Jackson Personality Inventory manual.* Research Psychologists Press, Port Huron, MI.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7 user's reference guide*. Chicago: Scientific Software International, Inc.

Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior, 16*, 215-224.

Kernberg, O. F. (1975). *Borderline conditions and pathological narcissism*. Northvale, NJ: Jason Aronson Inc.

Kernberg, O. F. (1996). A psychoanalytic theory of personality disorders. In J. F. Clarkin & M. F. Lenzenweger (Eds.), *Major theories of personality disorder* (pp. 106-140). New York: Guilford.

Kernberg, O. F., & Clarkin, J. F. (1995). *The Inventory of Personality Organization*. White Plains, NY: The New York Hospital-Cornell Medical Center.

Kohut, H. (1968). The psychoanalytic treatment of narcissistic personality disorders. *Psychoanalytic Study of the Child, 23*, 86-113.

Larsen, R. J., Diener, E., & Emmons, R. A. (1986). Affect intensity and reactions to daily life events. *Journal of Personality and Social Psychology, 51*, 803-814.

Leichsenring, F. (1999). Development and first results of the Borderline Personality Inventory: A self-report instrument for assessing borderline personality organization. *Journal of Personality Assessment, 73*, 45-63.

Lenzenweger, M. F. (1999). Stability and change in personality disorder features: The longitudinal study of personality disorders. *Archives of General Psychiatry, 56*, 1009-1015.

Lenzenweger, M. F., Clarkin, J. F., Kernberg, O. F., & Foelsch, P. A. (2001). The Inventory of Personality Organization: Psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a nonclinical sample. *Psychological Assessment, 13*, 577-591.

Lenzenweger, M. F., Lane, M. C., Loranger, A. W., & Kessler, R. C. (2007). DSM-IV personality disorders in the National Comorbidity Survey replication. *Biological Psychiatry, 62*, 553-564.

Levine, D., Marziali, E., & Hood, J. (1997). Emotion processing in borderline personality disorders. *Journal of Nervous and Mental Disease*, *185*, 240-246.

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York: The Guilford Press.

Loranger, A. W. (1999). *International personality disorder examination*. Lutz, FL: Psychological Assessment Resources, Inc.

Marsh, H. W. (1993). Self-esteem stability and responses to the Stability of Self Scale. *Journal of Research in Personality, 27*, 253-269.

Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling, 11*, 60-72.

Mennin, D. S., & Heimberg, R. G. (2000). The impact of comorbid mood and personality disorders in the cognitive-behavioral treatment of panic disorder. *Clinical Psychology Review, 20*, 339-357.

Millon, T. (1991). Classification in psychopathology: Rationale, alternatives, and standards. *Journal of Abnormal Psychology, 100*, 245-261.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45*, 19-30.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5^{th} Ed.). Los Angeles, CA: Authors.

Normandin, L., Sabourin, S., Diguer, L., Dupont, G., Poitras, K., Foelsch, P., et al. (2002). Évaluation de la validité théorique de l'Inventaire de l'organisation de la personnalité. *Revue Canadienne des Sciences du Comportement, 34*, 59-65.

Nurnberg, H. G., Raskin, M., Levine, P. E., Pollack, S., Siegel, O., & Prince, R. (1991). The comorbidity of borderline personality disorder and other *DSM-III-R* axis II personality disorders. *American Journal of Psychiatry, 148*, 1371-1377.

Oldham, J., Clarkin, J., Appelbaum, A., Carr, A., Kernberg, P., Lotterman, A., et al. (1985). A self-report instrument for borderline personality organization. In T. H. McGlashan (Ed.), *The borderline: Current empirical research* (pp. 3-18). Washington, DC: American Psychiatric Press, Inc.

Perry, J. C., & Ianni, F. F. (1998). Observer-rated measures of defense mechanisms. *Journal of Personality, 66*, 993-1024.

Rafaeli-Mor, E., & Steinberg, J. (2002). Self-complexity and well-being: A review and research synthesis. *Personality and Social Psychology Review, 6*, 31-58.

Rathus, J. H., & Miller, A. L. (1995). Life problems inventory. Unpublished manuscript. New York: Montefiore Medical Center/Einstein College of Medicine.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: Princeton University Press.

Ruuttu, T., Pelkonen, M., Holi, M., Karlsson, L., Kiviruusu, O., Heilä, H., Tuisku, V., et al. (2006). Psychometric properties of the Defense Style Questionnaire (DSQ-40) in adolescents. *Journal of Nervous and Mental Disease, 194*, 98-105.

Samuels, J., Eaton, W. W., Bienvenu, O. J., III, Brown, C. H., Costa, P. T., Jr., & Nestadt, G. (2002). Prevalence and correlates of personality disorders in a community sample. *British Journal of Psychiatry, 180*, 536-542.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Sharma, S., Durvasula, S., & Dillon, W. R. (1989). Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *Journal of Marketing Research, 26*, 214-221.

Shaw, D. S., Wagner, E. F., Arnett, J., & Aber, M. S. (1992). The factor structure of the Reckless Behavior Questionnaire. *Journal of Youth and Adolescence, 21*, 305-323.

Shea, M. T., Widiger, T. A., & Klein, M. H. (1992). Comorbidity of personality disorders and depression: Implications for treatment. *Journal of Consulting and Clinical Psychology, 60*, 857-868.

Shedler, J., Mayman, M., & Manis, M. (1993). The *illusion* of mental health. *American Psychologist, 48*, 1117-1131.

Skodol, A. E., Gunderson, J. G., Pfohl, B., Widiger, T. A., Livesley, W. J., & Siever, L. J. (2002). The borderline diagnosis I: Psychopathology, comorbidity, and personality structure. *Biological Psychiatry, 51*, 936-950.

Skowron, E. A., & Friedlander, M. L. (1998). The Differentiation of Self Inventory: Development and initial validation. *Journal of Counseling Psychology, 45*, 235-246.

Skowron, E. A., Holmes, S. E., & Sabatelli, R. M. (2003). Deconstructing differentiation: Self regulation, interdependent relating, and well-being in adulthood. *Contemporary Family Therapy, 25*, 111-129.

Skowron, E. A., Wester, S. R., & Azen, R. (2004). Differentiation of self mediates college stress and adjustment. *Journal of Counseling & Development, 82*, 69-78.

Spitzer, R. L. (2001). Values and assumptions in the development of DSM-III and DSM-III-R: An insider's perspective and a belated response to Sadler, Hulgus, and Agich's "On values in recent American psychiatric classification". *Journal of Nervous and Mental Disease, 189*, 351-359.

Stone, M. H. (1993). Long-term outcome in personality disorders. *British Journal of Psychiatry, 162*, 299-313.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*, 297-303.

Torgersen, S., Kringlen, E., & Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archives of General Psychiatry, 58*, 590-596.

Trijsburg, R. W., Van T'Spijker, A., Van, H. L., Hesselink, A. J., & Duivenvoorden, H. J. (2000). Measuring overall defensive functioning with the Defense Style Questionnaire: A comparison of different scoring methods. *Journal of Nervous and Mental Disease, 188*, 432-439.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70.

Wagner, M. K. (2001). Behavioral characteristics related to substance abuse and risk-taking, sensation-seeking, anxiety sensitivity, and self-reinforcement. *Addictive Behaviors, 26*, 115-120.

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*, 1063-1070.

Watson, D. C. (2002). Predicting psychiatric symptomatology with the Defense Style Questionnaire-40. *International Journal of Stress Management, 9*, 275-287.

Wei, M., Vogel, D. L., Ku, T.-Y., & Zakalik, R. A. (2005). Adult attachment, affect regulation, negative mood, and interpersonal problems: The mediating roles of emotional reactivity and emotional cutoff. *Journal of Counseling Psychology, 52*, 14-24.

Widiger, T. A., & Frances, A. J. (1989). Epidemiology, diagnosis, and comorbidity of borderline personality disorder. In A. Tasman, R. E. Hales, & A. J. Frances (Eds.), *American Psychiatric Press Review of Psychiatry: Vol. 8* (pp. 8-24). Washington, DC: American Psychiatric Press.

Yen, S., Zlotnick, C., & Costello, E. (2002). Affect regulation in women with borderline personality disorder traits. *Journal of Nervous and Mental Disease, 190*, 693-696.

Zanarini, M. C., Frankenburg, F. R., Dubo, E. D., Sickel, A. E., Trikha, A., Levin, A., et al. (1998). Axis I comorbidity of borderline personality disorder. *American Journal of Psychiatry*, *155*, 1733-1739.

Zanarini, M. C., Frankenburg, F. R., Vujanovic, A. A., Hennen, J., Reich, D. B., & Silk, K. R. (2004). Axis II comorbidity of borderline personality disorder: Description of 6-year course and prediction to time-to-remission. *Acta Psychiatrica Scandinavica*, *110*, 416-420.

Ziller, R. C., Martell, R. T., & Morrison, R. H. (1977). Social insulation, self-complexity, and social attraction: A theory chain. *Journal of Research in Personality*, *11*, 398-415.

Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of *DSM-IV* personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, *162*, 1911-1918.