7

Methodological Issues in the Assessment of the Affective Component of Subjective Well-Being

Ulrich Schimmack

One important area of positive psychology examines happiness and subjective well-being (SWB; Diener, 2000). Diener (1984) noted that SWB has an affective and a cognitive component. The cognitive component of SWB is assessed with life satisfaction judgments (Diener, Emmons, Larsen, & Griffin, 1985). The affective component assesses the amount of pleasant and unpleasant experiences in people's lives (e.g., Schimmack, Diener, & Oishi, 2002).

Early on, SWB researchers noted the limitations of research programs that focused exclusively on negative states such as depression and anxiety. In an influential article, Diener (1984) proposed that SWB is more than the absence of negative affect (NA). Although low levels of NA are important for SWB, high levels of positive affect (PA) are also important. Diener's (1984) conception of SWB in terms of high PA and low NA implies that PA and NA are separable components of SWB. In other words, measures of PA show discriminant validity from measures of NA. If PA and NA were bipolar opposites, then minimization of NA would also maximize PA. In contrast, the distinction of PA and NA as separate components of SWB implies that PA and NA have different causes and consequences. The conceptualization of PA and NA as separate dimensions of SWB created a heated debate in the affect literature (Diener, 1999). This chapter gives a brief overview of the debate with a focus on the methodological issues that have fuelled the debate.

Positive Affect and Negative Affect

In the early 1980s, several independent lines of research converged on the view that PA and NA are separable (i.e., not bipolar opposite) dimensions. Factor analyses of self-reported affects often resulted in two factors, with pleasant items loading on one factor and unpleasant items loading on the other factor. Personality researchers noted that PA and NA are related to different personality traits. Whereas extraversion is a stronger predictor of PA than NA, neuroticism is a stronger predictor of NA than PA (Costa & McCrae, 1980). These findings provided the foundation for the distinction between PA and

NA (Diener, 1984; see Diener & Lucas, 1999, for a review).

Discriminant Validity of PA and NA: A Method Artifact?

Green, Goldman, and Salovey (1993) presented the first serious challenge to the view of PA and NA as separable dimensions. They proposed that evidence of discriminant validity of PA and NA scales is a method artifact. In particular, an acquiescence bias would bias the correlation between observed ratings of PA and NA in a positive direction because respondents with an acquiescence bias agree to PA items and NA items independent of their actual level of SWB. Green et al. (1993) proposed a multiformat approach to control for the influence of response styles. Concretely, PA and NA were assessed with multiple response formats such as Likert scales and adjective checklists. The data were analyzed with structural equation models. Structural equation models essentially decomposed the variance in each item into two variance components: variance that was shared with other response formats, and variance that was unique to a particular response format. The first variance component is assumed to reflect valid variance, whereas the second component reflects both random measurement error and systematic measurement error due to response styles. The model also allowed for correlated error variances between measures that used the same response format. The magnitude of these correlations reflects the influence of systematic measurement error. Green et al.'s (1993) multiformat studies yielded format-independent correlations between PA and NA that approached -1 ($\sim -.90$), indicating that respondents' level of PA was highly inversely related to their level of NA. This finding has been widely cited as evidence against the assumption that PA and NA are separable dimensions.

Only a few years after Green et al.'s (1993) challenge, Diener, Smith, and Fujita (1995) reexamined the relation between PA and NA by means of a multitrait multimethod analysis. PA daily diary reports and retrospective reports is relatively free of memory biases. Second, self-reports and informant reports are made by different individuals. Thus, shared variance between these measures is free of response styles. Importantly, the use of informant report is superior to Green et al.'s multiformat procedure, which assumes that response styles are unique to a particular response format. Thus, the multiformat procedure fails to control for response styles that are common to all response formats. For example, people who respond in a social desirable manner on one format are also likely to do so on other formats. Using structural equation modeling, Diener et al. (1995) found that the method-free correlation between PA and NA was $-.44$. This finding contradicts Green et al.'s (1993) conclusion that PA and NA lack discriminant validity once random and systematic measurement error are controlled. Subsequently, I review attempts to reconcile the divergent results of Green et al. (1993) and Diener et al. (1995).

Cultural Variation in the Relation Between PA and NA

Bagozzi, Wong, and Yi (1999) proposed that the relation between PA and NA varies across samples. They presented empirical support for this hypothesis in a cross-cultural study. The correlations between PA and NA were less negative in East Asian cultures than in Western cultures. The authors attributed this finding to dialectic thinking in Asian philosophy. In cultures with dialectic thinking, opposite affective states are seen as compatible, which leads to weaker negative correlations between PA and NA. Schimmack, Oishi, and Diener (2002) extended Bagozzi et al.'s findings in a larger cross-cultural study of 38 nations. The data revealed that weaker negative correlations were indeed unique to East Asian cultures and not a general characteristic of non-Western cultures. However, Schimmack, Oishi, et al. (2002) also noted that even nondialectic samples had rather modest negative correlations compared to Green et al.'s (1993) findings. In addition, it has to be

and NA were assessed with three different methods, namely (a) retrospective self-reports, (b) daily diaries, and (c) informant reports. This approach has several advantages over the multiformat procedure used by Green et al. (1993). First, daily diary data control for memory biases in retrospective self-reports and informant reports. Thus, any variance that is shared between

reports with a single method. Thus, the finding of discriminant validity in these studies could be due to response styles.

Item Content of PA and NA Measures

Watson and Tellegen (1999) provided an alternative explanation for discrepant results regarding the discriminant validity of PA and NA (see also Tellegen, Watson, & Clark, 1999). Accordingly, evidence for discriminant validity of PA and NA depends on the item content of PA and NA scales. Green et al.'s (1993) challenge to the discriminant validity of PA and NA was limited to scales that equated PA with happiness and NA with sadness. In contrast, multimethod and multiformat studies with other PA and NA scales showed clear evidence of discriminant validity (Green et al., 1993; Lucas, Diener, & Suh, 1996). However, Watson and Tellegen's (1999) explanation overlooked that Diener et al.'s (1995) study included happiness and sadness as components of PA and NA. Furthermore, Diener et al. (1995) reported the method-free correlation between happiness and sadness. This correlation ($r = -.47$) also supported discriminant validity and contradicts Green et al.'s (1993) finding of much stronger correlations that approach -1. Furthermore, Diener et al.'s (1995) data showed that the correlation between happiness and sadness was not dramatically different from correlations between happiness and other unpleasant emotions (r $-.33$ to $-.37$). This finding is consistent with other findings that correlations between PA and NA measures do not vary dramatically with scale content (Schimmack, 2003; Schimmack, Oishi, et al., 2002; Watson, 1988; Zelenski & Larsen, 2000). Thus, item content of PA and NA scales alone is unable to reconcile the findings of Green et al. (1993) and Diener et al. (1995).

noted that Green et al. and Diener et al. (1995) both used North American student samples. Thus, cultural variations in the relation between PA and NA are theoretically important, but they cannot explain the inconsistent findings across North American samples. One weakness of Bagozzi et al.'s (1999) and Schimmack, Oishi, et al.'s (2002) studies was the reliance on self-

1999). Meddis (1972) distinguished asymmetrical and symmetrical formats. Asymmetrical formats have a single rejection category and multiple acceptance categories (e.g., not at all, a little, a lot). Symmetrical formats have an equal number of rejection and acceptance categories (e.g., strongly disagree, disagree, neither, agree, strongly agree). Symmetrical formats imply a bipolar construct because it is meaningless to distinguish degrees of absence. For example, it does not make sense to be slightly not happy or strongly not happy. Thus, faced with a symmetrical scale, respondents assume that multiple rejection categories should be used to indicate the extent to which the opposite construct is present. For example, slightly not happy is translated into slightly sad and strongly not happy is translated into strongly sad. Three of the four response formats in Green et al.'s (1993) study were symmetrical formats that respondents were likely to interpret as bipolar scales (cf. Russell & Carroll, 1999). In contrast, Diener et al. (1995) used asymmetrical formats for the retrospective self-reports and informant reports as well as the daily diary data. Thus, the nature of the response formats provides a simple explanation for the divergent findings.

Schimmack, Bockenholt, et al. (2002) tested this explanation in a series of studies. One study demonstrated that the correlation between PA and NA could be predicted on the basis of an independent measure of respondents' interpretation of the response format. Another study demonstrated that PA and NA were only moderately negative correlated when multiple asymmetrical formats were used. Schimmack, Bockenholt, et al. (2002) also demonstrated that Green et al.'s (1993) explanation for their strong negative correlation was incorrect. The authors had attributed their finding to the ability of studies that vary response formats to control systematic measurement error. A reexamination

Symmetrical Versus Asymmetrical Response Formats

Schimmack, Bockenholt, and Reisenzein (2002) offered another explanation for the discrepancy between the findings of Green et al. (1993) and Diener et al. (1995). They proposed that the authors used different response formats, which produced different correlations between PA and NA. Numerous articles in the affect literature had demonstrated that the correlation between PA and NA varies with the nature of a response format (e.g., Meddis, 1972; Russell & Carroll, of their own data revealed that none of their three rating scales were significantly influenced by systematic measurement error. Thus, each of their symmetric response formats alone would have produced evidence for bipolarity after controlling for random measurement error alone.

Schimmack, Bockenholt, et al.'s (2002) findings provide evidence that the nature of response formats explains the discrepant findings regarding PA and NA in the literature. This finding leads to the important question of which response formats provide valid information about the relation between PA and NA. I argue that

asymmetrical formats with a single rejection category are superior to symmetrical formats. The reason is that symmetrical formats are interpreted as bipolar scales, which is inappropriate for investigations of the relation between PA and NA. If the negative pole of a PA measure is defined as high levels of NA and the negative pole of a NA measure is defined as high levels of PA, then the two measures must be inversely related. Indeed, any correlation lower than −1 must be due to measurement error, but cannot be interpreted as evidence for discriminant validity of PA and NA. Thus, an assessment of PA and NA with asymmetrical scales is required to study the relation between variation in PA and NA.

Pearson Correlation Versus Polychoric Correlations

Although response formats play an important role in determining the correlation between PA and NA, other methodological factors may also influence the relation. Some researchers have discussed the possibility that the Pearson correlation is an inappropriate statistical test (Eid, Notz, Schwenkmezger, & Steyer, 1994; Tellegen et al., 1999). The reason is that the Pearson correlation assumes that observed variables are normally distributed. If variables are not normally distributed, Pearson correlations provide attenuated estimates of the relation between two variables.

The assumption of normal distributions is typically violated for state measures of NA. Most people report not feeling any negative affect, which leads to positively skewed item distributions (e.g., Schimmack, 2003). Polychoric correlation coefficients rectify the problem of nonnormal distributions by mapping nonnormally distributed observed variables onto a normally distributed latent variable. A polychoric correlation coefficient estimates the association between two latent normally distributed variables for observed variables that do not fulfill the normal-distribution assumption of Pearson correlations.

negative correlation ($r = -.91$). This finding suggests that evidence for discriminant validity of PA and NA in Diener et al.'s (1995) study may have been due to an inappropriate use of the Pearson correlation with nonnormally distributed variables. However, other findings in the literature cast doubt on this interpretation. First, Tellegen et al. (1999) did not report the Pearson correlation for the same data set. Thus, it is unknown whether a Pearson correlation would have produced substantially different results. Eid et al. (1994) compared factor structures based on Pearson correlations and polychoric correlations and found that the "differences between both models are small" (p. 211). Watson and Tellegen (1999) reported that the Pearson correlation between single items of happiness and sadness was only slightly less negative ($r = -.48$) than the polychoric correlation ($r = -.57$). Furthermore, Diener et al. (1995) examined trait affect. Unlike state measures of NA, trait measures of NA tend to conform better to the assumption of normal distributions. Thus, it is likely that Diener et al. (1995) would have obtained similar findings if they had used polychoric correlations.

So far, nobody has compared Pearson and polychoric correlations with trait measures of PA and NA. For this purpose, I analyzed two new data sets. One study asked 517 students how often they experienced happiness and sadness in general. The response format was a 7-point scale ranging from 1 (almost never) to 7 (nearly always). The second study asked 819 students how often they felt happiness and sadness in the past month (30 days). The response format was an 8-point scale ranging from 0 (never, 0%) to 7 (always, 100%). Each affect was assessed with three items to control for random measurement error. The items were happy, cheerful, and joyful for PA and sad, depressed, and blue for NA. The correlation between PA and NA was tested using structural equation modeling. Each model had two factors and the three items of each construct loaded on one of the two factors.

Tellegen et al. (1999) used polychoric correlations to estimate the correlation between happiness and sadness. They also used structural equation modeling to control random measurement error. The polychoric correlation between happiness and sadness was −.77, which is a considerably more negative correlation than the Pearson correlation found by Diener et al. (1995). A model that further controlled for acquiescence biases produced an even stronger

Separate models were tested with Pearson and with polychoric correlation coefficients. Models with polychoric correlation coefficients require large sample sizes and the number of participants increases exponentially with the number of response categories. To obtain robust estimates of polychoric correlations, models were tested with the original 7-point and 8-point response formats and with data that reduced the number of response categories to three by

combining neighboring response categories (see Eid et al., 1994). The estimated correlations with all response categories and combined categories were virtually identical. Thus, the results with the full response format are reported. Polychoric correlation coefficients were nearly equivalent to Pearson correlations. For affect in general, the Pearson correlation was $r = -.37$ and the polychoric correlation was $r = -.38$. For affect in the past month, the Pearson correlation was $r = -.46$ and the polychoric correlation was $r = -.45$. The reason for this small effect of the statistical coefficient can be found in the distribution of the various measures. Neither the happiness scale (skewness = −0.33, −0.15; general and past month, respectively) nor the sadness scale (skewness = 0.41, 0.68) deviated dramatically from normality.

To examine the influence of the type of correlation coefficient on analyses of state NA and PA, I examined the correlation between ratings of momentary PA and NA in a large student sample ($N = 710$; see Schimmack & Reisenzein, 2002). PA was assessed with the items pleasant, good, and positive. NA was assessed with the items unpleasant, bad, and negative. Responses were made on a 4-point scale ranging from 0 (not at all) to 3 (strongly). Consistent with expectations, the NA scale was positively skewed (skewness = 1.40), whereas PA was approximately normally distributed (skewness = −0.42). The data were analyzed with the same structural equation model as the trait data. Models were tested with the full 4-point scale and a reduced 3-point scale that combined the least frequently used category with the neighbor category. This was 0 for the PA items and 3 for the NA items. Both models provided virtually identical results and the results for the 4-point scale are reported. The Pearson correlation was estimated to be $r = -.61$, whereas the polychoric correlation was estimated to be $r = -.66$. This finding shows that Pearson correlations are likely to attenuate the correlation between PA and NA

out that response styles may produce systematic measurement error in studies that aggregate state measures to obtain trait measures (Schimmack, 2003; Schimmack, Bockenholt, et al., 2002; Watson & Tellegen, 2002). The reason is that aggregation increases the amount of systematic variance, which applies equally to systematic trait variance and systematic measurement error. In support of this hypothesis, Schimmack, Bockenholt, et al. (2002) demonstrated that Diener et al.'s (1995) daily diary data systematically underestimated the correlation between PA and NA in comparison to the retrospective self-reports and informant reports. Schimmack, Bockenholt, et al. estimated that aggregated response styles change correlations by about .2. A similar estimate was obtained by Schimmack (2003) using a different measure of response styles. Participants rated a set of facial expressions of emotions using the same response format. Assuming that ratings of others' emotions are independent of individuals' own levels of PA and NA, correlations between self-ratings and rating of facial expressions reflect the influence of response styles. Controlling for response styles, the correlation between aggregated momentary rating of PA and NA changed from .02 to −.15. Watson and Tellegen (2002) report similar discrepancies between uncorrected and corrected correlations of aggregated PA and NA measures. The influence of systematic measurement error can explain the near-zero correlations between happiness and sadness in studies that used aggregated momentary ratings (e.g., Schimmack, 2003; Zelenski & Larsen, 2000). The true correlations are likely to be negative and similar to those obtained in Diener et al.'s (1995) multimethod study. However, even these negative correlations would still support the discriminant validity of PA and NA.

Variation in PA-NA Correlations: A Mini Meta-Analysis

in analyses of state affect due to the skewed distribution of NA items. However, the effect is modest and cannot explain the discriminant validity of PA and NA.

Aggregation Versus Retrospective Trait Measures

As noted earlier, response styles seem to have a negligible effect on the correlation between PA and NA. However, recent publications pointed

The review of methodological issues in the measurement of PA and NA has uncovered several factors that influence the correlation between PA and NA. For a quantitative assessment of the contribution of these factors to the variation in PA-NA correlation, I conducted a meta-analysis. The point of the meta-analysis was not to include all studies of the correlation between PA and NA. Rather, the focus was on studies that varied across the different factors that may contribute to variation in the

TABLE 7.1 Independent Variables and PA-NA Correlations in Studies Used for the Mini Meta-Analysis

Study	SM/MM	Format	State/Trait	Statistic	Aggregation	Correlation
Green et al. (1993) Study 1	1	1	1	0	0	−.85
Green et al. (1993) Study 2	1	1	1	0	0	−.91
Diener et al. (1995)	1	0	0	0	0	−.47
Barrett & Russell (1998) Study 1	1	1	1	0	0	−.93
Barrett & Russell (1998) Study 2	1	1	1	0	0	−.93
Russell & Carroll (1999) 1	0	0	1	0	0	−.46
Russell & Carroll (1999) 2	0	0	1	0	0	−.51
Russell & Carroll (1999) 3	0	0	1	0	0	−.56
Russell & Carroll (1999) 4	0	1	1	0	0	−.79
Tellegen et al. (1999)	0	0	1	1	0	−.77
Tellegen et al. (1999)	0	0	1	1	0	−.91
Watson & Tellegen (1999)	0	0	1	0	0	−.48
Watson & Tellegen (1999)	0	0	1	1	0	−.57
Zelenski & Larsen (2000) WS	0	0	1	0	0	.30
Zelenski & Larsen (2000) AGG	0	0	0	0	0	−.13
Schimmack, Bockenholt, et al. (2002)	1	0	1	0	0	−.52
Schimmack et al. (2002) NA	0	0	0	0	0	−.34
Schimmack et al. (2002) A	0	0	0	0	1	−.13
Schimmack (2003) WS	0	0	1	0	0	−.22
Schimmack (2003) AGG	0	0	0	0	1	−.02
Schimmack (2004) General	0	0	0	0	0	−.37
Schimmack (2004) General	0	0	0	1	0	−.38
Schimmack (2004) Month	0	0	0	0	0	−.46
Schimmack (2004) Month	0	0	0	1	0	−.45
Schimmack & Reisenzein (2002)	0	0	1	0	0	−.61
Schimmack & Reisenzein (2002)	0	0	1	1	0	−.66

Note: SM/MM = single method/multimethod study; NAs not Asian dialectic; A = Asian dialectic; WS = within-subject; AGG = aggregated states.

PA-NA correlation. Table 7.1 lists all the studies that were included in the analyses. It also lists the variables that were used as predictors of variation in PA-NA correlations, namely single-method versus multimethod assessment, symmetric versus asymmetric response format, state versus trait affect, Pearson versus polychoric correlation coefficient, and one-time versus aggregate assessment. The PA-NA correlations show a large range of variability that would support models of near independence ($r = -.02$) to approximate bipolarity ($r = -.93$). Table 7.2 shows the correlation among predictors and their correlation with the PA-NA correlation coefficients. First, it is remarkable that most of the predictors are only weakly related to each other, indicating that the meta-analysis covers a range of studies with different characteristics. Second, the simple correlations indicated that response format is the strongest

TABLE 7.2 Correlations Between Method Factors and PA-NA Correlations

	Method	Format	State/Trait	Statistic	Aggregation	Correlation
Method	—					
Format	.66*	—				
State/trait	.23	.38	—			
Statistic	−.28	−.25	−.04	—		
Aggregate	−.17	−.15	−.39	−.15	—	
Correlation	−.58*	−.74*	−.63*	−.11	.52*	—

*$p < .05$

predictor of PA-NA correlations. This finding is consistent with studies that directly examined the effect of response format on PA-NA correlation (Meddis, 1972; Russell & Carroll, 1999; Schimmack, Bockenholt, et al., 2002). A regression analysis with format and method as predictors revealed that format predicted unique variance in PA-NA correlations ($\beta = -.64$), whereas method had no significant effect ($\beta = -.16$), $t < 1$. This finding confirms Schimmack, Bockenholt, et al.'s conclusion that multimethod studies produce correlations similar to those of single-format studies if the nature of the formats is the same.

When PA-NA correlations were regressed onto format, state versus trait, type of correlation coefficient, and aggregation, all predictors contributed significantly to the prediction of PA-NA correlations. The model explained 83% of the variance in PA-NA correlations across studies.

$$\begin{aligned}\text{PA-NA CORR} = &-.33 - .41\ \text{format} \\ &+ .25\ \text{aggregation} \\ &- .15\ \text{state/trait} \\ &- .15\ \text{statistic}\end{aligned}$$

This formula implies that the correlation with asymmetrical formats is −.33, whereas the correlation with symmetrical formats is −.74. As noted earlier, symmetrical formats are unsuitable for the examination of the relation between PA and NA. Thus, format should be set to zero. With format set to zero, the formula is

$$\begin{aligned}\text{PA-NA CORR} = &-.33 + .25\ \text{aggregation} \\ &- .15\ \text{state/trait} \\ &- .15\ \text{statistic}\end{aligned}$$

As noted earlier, aggregation attenuates the true correlation due to the influences of systematic measurement error. Thus, aggregation should be set to zero, which simplifies the formula further.

of $r = -.22$, which was predicted to be $r = -.47$. As Schimmack's (2003) finding is based on single item indicators, the attenuated correlation is likely due to random measurement error.

In sum, this review has revealed several factors that contribute to the inconsistent findings of the discriminant validity of PA and NA. The strongest effect was obtained for the type of response format used. The strongest challenge to discriminant validity stems from studies with symmetrical response formats. These formats are unsuitable for the investigation of the relation between PA and NA because respondents interpret them as bipolar scales. On the other hand, studies with aggregated state measures overestimate discriminant validity due to response styles that become more notable with aggregation. The use of polychoric versus Pearson correlation has a small effect on the results that is negligible for tests of discriminant validity. Furthermore, the effect is likely to be stronger for state measures than for trait measures. State PA and NA show slightly more negative correlations than trait measures, but the difference is also quite small. To conclude, the present review reveals that Diener et al.'s (1995) multimethod study provided valid information about the discriminant validity of PA and NA, whereas Green et al.'s (1993) challenge was due to the reliance on inappropriate symmetrical formats.

The present analyses also do not confirm Tellegen et al.'s (1999) finding that state measures of PA and NA are bipolar opposites with happiness and sadness as PA and NA items. The meta-analysis predicts a polychoric correlation of $r = -.63$, which corresponds to the analysis of Schimmack and Reisenzein's (2002) data ($r = -.66$), but is considerably weaker than Tellegen et al.'s (1999) estimate ($r = -.91$). Tellegen et al.'s (1999) study is unique in several ways, but each of these methodological factors alone is unable to explain their results. First,

PA-NA CORR = −.33 − .15 state/trait
− .15 statistic

The final formula suggests that PA-NA correlation can range from −.33 to −.63, with stronger correlations for studies of states than traits and studies that use polychoric rather than Pearson correlations. An examination of the residuals showed no major discrepancies between actual and predicted scores. The biggest discrepancy was found for Schimmack's (2003) finding of a within-subject PA-NA correlation

Tellegen et al. (1999) examined state affect, whereas Diener et al. (1995) examined trait affect. However, Schimmack (2001, 2005) demonstrated that even state measures of pleasure and displeasure show discriminant validity. Second, Tellegen et al. (1999) used polychoric correlation coefficients. However, other studies with polychoric correlation coefficients still support discriminant validity, although the negative correlations are quite strong (Eid et al., 1994). Third, Tellegen et al. (1999) controlled response styles. However, Schimmack, Bockenholt, et al.

(2002) found no systematic effect of response styles on one-time assessments of PA and NA, and Watson and Tellegen (2002) also argued that response styles do not have a strong effect on single assessments of PA and NA. With correlations in the range of −.6 to −1, it becomes increasingly difficult to examine discriminant validity. As noted by Russell and Carroll (1999), evidence for discriminant validity rests on the assumption that researchers have successfully eliminated and controlled any source of measurement error. As this assumption is impossible to prove, it remains possible to defend bipolar models of state PA and NA. Fortunately, studies that go beyond psychometric analyses overcome this problem. These studies are reviewed next.

Mixed Feelings

Russell and Carroll (1999) argued that discriminant validity does not necessarily imply that PA and NA are separate affects. It is important to note, however, that their argument is limited to state measures of PA and NA with asymmetrical scales. Russell and Carroll's (1999) argument is best explained with a visual illustration. Figure 7.1 shows asymmetrical PA and NA scales that ranges from 0 (not at all) to 3 (extremely). The figure is a contingency table of the different levels of PA and NA. The hypothetical data pattern in figure 7.1 assumes that PA and NA are mutually exclusive. If somebody feels happy, he or she cannot feel sad at the same time and vice versa. Thus for any value greater than zero on PA, the value for NA must be zero, and for any value greater than zero on NA, the value for PA must be zero.

Mutually exclusive states of PA and NA do not produce a linear relationship between PA and NA. Indeed, Pearson correlations between PA and NA can range from close to zero up to −1, depending on the distribution of PA and NA (see Schimmack, 2001, for an illustration). If a study samples mostly weak affective experiences, the linear correlation is closer to zero. If a study samples mostly intense affective experiences of both valences, the linear correlation becomes increasingly negative. Russell and Carroll's (1999) model of mutually exclusive states of PA and NA is compatible with traditional notions of discriminant validity. The reason is that, in figure 7.1, it is impossible to predict PA values from NA values and vice versa. For example, when NA is 0, PA can range from 0 to 3, and when PA is 0, NA can range from 0 to 3. As a result, asymmetrical scales of PA and NA provide unique information about an individual's SWB, even when PA and NA are mutually exclusive.

However, Russell and Carroll's (1999) model suggests that it is unnecessary to measure PA and NA separately because a single bipolar measure that covers the full range of PA and NA provides the same information as separate measures of PA and NA. For the hypothetical data pattern in figure 7.1, a bipolar measure that ranges from −3 to +3 allows perfect prediction of the unipolar PA and NA scores in figure 7.1.

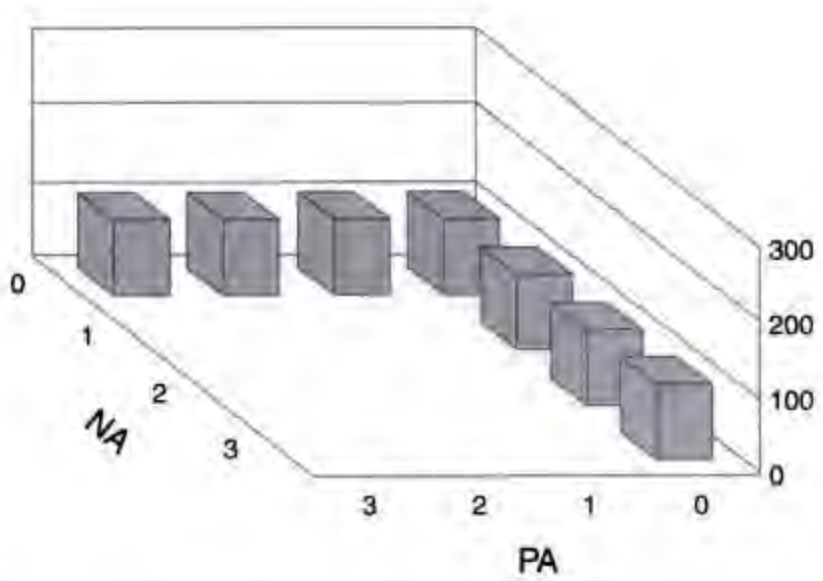

Figure 7.1 Russell and Carroll's model of mutually exclusive states of PA and NA.

For example, if the score on the bipolar measure is −2, then the NA score is 2 and the PA score is 0. If the bipolar score is +2, then PA is 2 and NA is 0. Thus, Russell and Carroll's (1999) model still poses a serious challenge to the notion of PA and NA as separate dimensions of SWB. However, the challenge assumes that state PA and NA are indeed mutually exclusive. Only a few studies have tested this hypothesis.

The first empirical test of Russell and Carroll's (1999) bipolar model was carried out by Diener and Iran-Nejad (1986; see also Beebe-Center, 1932, for a review of earlier, introspective studies of this issue). The authors asked participants to rate their affective experience on a set of affect items during everyday emotional experiences. Diener and Iran-Nejad (1986) found that high levels of PA and NA are mutually exclusive, but low levels of PA and NA are not (see figure 7.2 for an illustration). This finding undermines the assumption that state PA and NA are mutually exclusive.

Russell and Carroll (1999) argued that data patterns like Diener and Iran-Nejad's (1986) are methodological artifacts due to the selection of inappropriate items, inappropriate response formats, random measurement error, and response styles. However, several more recent studies have carefully controlled for potential method artifacts and still found evidence for mixed feelings; that is, concurrent reports of PA and NA (Hemenover & Schimmack, 2004; Hunter, Schellenberg, & Schimmack, 2006; Larsen, McGraw, & Cacioppo, 2001; Larsen, McGraw, Mellers, & Cacioppo, 2004; Schimmack, 2001, 2005; Schimmack & Colcombe, in press).

To demonstrate the validity of reports of mixed feelings, these studies used an experimental approach with a control condition. The common assumption of these studies is that reports of mixed feelings should vary across situations. Only ambivalent situations that can be appraised in a positive or negative manner should elicit mixed feelings. As a result, models with separate dimensions of PA and NA and Russell and Carroll's (1999) model make different predictions about reports of mixed feelings in different situations. Russell and Carroll's model assumes that reports of mixed feelings are due to method artifacts. Thus, the model predicts that reports of mixed feelings are the same across different situations. In contrast, models with separate dimensions of PA and NA assume that reports of mixed feelings are more frequent and more intense during ambivalent situations than during unambiguous situations.

Schimmack (2001) tested these predictions in a study of changes in affective experiences. A model with mutually exclusive states of PA and NA has to explain how affect changes when people experience one affect (e.g., PA) and are confronted with a stimulus of the opposite valence (e.g., a negative stimulus). To illustrate, imagine somebody walking along a beach in a good mood. At this moment, the individual experiences PA and NA is absent. Then the individual notices trash on the beach. How does this event influence the individual's affective experience? A model with mutually exclusive experiences of PA and NA would have to assume one of two scenarios. First, the trash fully ruins the pleasant experience of being on a sunny beach.

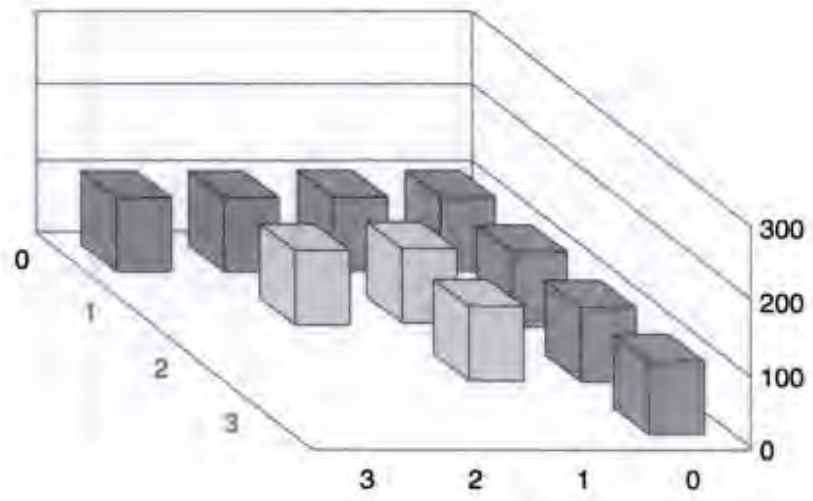

Figure 7.2 Schematic display of Diener and Iran-Nejad's (1986) findings.

Accordingly, the individual now feels NA without PA. Second, the trash merely reduces the intensity of PA, but NA remains absent. In contrast, a model with separate dimensions of PA and NA allows for the possibility that the trash induces feelings of NA without fully suppressing the initial feeling of PA, although its intensity may be reduced.

The different predictions were tested in an experiment with a mild negative mood induction. Before the mood induction, most participants reported moderate levels of PA and the absence of NA. After the mood induction, PA decreased in intensity and NA increased. Because participants started with a moderate level of PA, they still experienced PA after the mood induction. As a result, more participants experienced mixed feelings of PA and NA after the mood induction than before the mood induction. Given the significant difference in the frequency and intensity of mixed feelings between the two conditions, the results cannot be attributed to measurement artifacts. Thus, Schimmack's (2001) results are consistent with Diener and Iran-Nejad's (1986) model of separate dimensions of PA and NA and inconsistent with Russell and Carroll's (1999) model of mutually exclusive experiences of PA and NA.

Schimmack (2005) replicated and extended these findings. One extension was the manipulation of the order and the time lag between the pleasure and displeasure items. Neither of these variables had an effect on the intensity of mixed feelings after the experiment, despite a large sample size ($N = 901$). This finding suggests that reports of mixed feelings were not due to changes in affective experiences over the course of the mood assessment.

Larsen, McGraw, and Cacioppo (2001) examined experiences of happiness and sadness in ordinary situations and in ambivalent situations

Larsen et al. (2004) found further evidence for mixed feelings with a gambling paradigm. Each gamble had two possible outcomes of the same valence (e.g., winning $5 or $12). The authors proposed that affective reactions are determined by the valence of the actual outcome (i.e., wins elicit PA, losses elicit NA) and the discrepancy with the alternative outcome (i.e., winning $5 is less desirable than winning $12). Some situations are unambiguously positive or negative (e.g., winning $12 in a gamble with $5 and $12 payoffs). Other situations are ambiguous because they are desirable in one way (winning $5), but undesirable in another way (not winning $12). Larsen et al. (2004) predicted and found that ambiguous outcomes elicited mixed feelings more frequently and intensely than unambiguous outcomes. Furthermore, the intensity of mixed feelings varied with the discrepancy between the obtained and not-obtained outcome. That is, the strongest mixed feelings were obtained for winning $5 and not winning $12, and for losing $5 and not losing $12.

A second study addressed the concern that participants experienced PA and NA in rapid alternations rather than concurrently. For this purpose, participants had to press separate keys for PA and NA as long as they experienced either affect. Rather than alternating their responses, participants pressed both keys for extended periods of time. This finding suggests that PA and NA are separate affects that can be experienced concurrently.

Whereas the previous studies focused on pleasure and displeasure and happiness and sadness, Hemenover and Schimmack (2004) examined the compatibility of two other affects, namely amusement and disgust. Participants watched a film clip with disgusting humor. Participants reported mixed feelings of disgust and amusement more frequently after than before the film clip.

that could elicit happiness and sadness, such as viewing a happy and sad movie (*Life Is Beautiful*) or leaving one's dorm on graduation day. The authors found that participants reported more mixed feelings of happiness and sadness during the ambivalent situations than in normal control situations.

Schimmack and Colcombe (in press) used an old paradigm by Kellogg (1915) to examine mixed feelings. Participants were confronted with presentations of two emotional pictures. Presentations of ambivalent picture pairs elicited more intense mixed feelings than presentations of unambiguous picture pairs.

Hunter et al. (2006) examined affective reactions to music. It is well known that tempo (fast vs. slow) and mode (major vs. minor) influence affective reactions to music in opposite ways. Fast music in major mode is happy, whereas slow music in minor mode is sad. Hunter et al. (2006) found that music with conflicting affective cues (e.g., fast minor or slow major) elicit more intense mixed feelings than music with consistent affective cues (e.g., fast major or slow minor).

In sum, these findings indicate that experiences of PA and NA are not mutually exclusive. As a result, it impossible to infer levels of PA and NA from a single bipolar score. For example,

a score of +1 on a bipolar scale could be due to +1 PA and 0 NA, +2 PA and +1 NA, or +3 PA and +2 NA. Thus, even studies of momentary experiences of PA and NA are more consistent with the view that PA and NA are separable dimensions of SWB (Cacioppo & Berntson, 1994; Diener & Iran-Nejad, 1986).

PA and NA Scales

The previous section did not distinguish between different measures of PA and NA because all measures of PA and NA show discriminant validity independent of item content. Furthermore, different measures of PA and NA also often show high convergent validity (Watson, 1988). Thus, it is likely that different PA or NA scales yield similar results. Nevertheless, some studies have revealed discrepant results with different measures of PA and NA. Most of these studies have focused on a comparison of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) with other measures of PA and NA. The PANAS scales differ from other scales in that they were developed on the basis of a psychometric model of affect that postulates two orthogonal dimensions. To obtain this objective, the scales include some items (e.g., alert, interested) that are uncommon on other measures, and they do not include items (e.g., happy, sad) that are common on other scales. These variations in item content can explain why PANAS scales sometimes produce different results than other scales.

Patrick and Lavoro (1997) examined affective reactions to pictures from the International Affective Picture System (IAPS). The IAPS contains a wide range of pictures from extremely positive to extremely negative ones. The authors hypothesized that positive pictures influence

Another relevant finding concerns the effect of weekdays versus weekends on PA. Typically, people report more pleasant experiences on weekends than on weekdays (e.g., Stone, Hedges, Neale, & Satin, 1985). In contrast, Clark and Watson (1988) found no differences in PANAS-PA between weekdays and weekends. Egloff, Tausch, Kohlmann, and Krohne (1995) directly compared PANAS-PA and another PA scale and found significantly effects of weekend versus weekday for the ordinary PA scale, but not for PANAS-PA. Once more, this finding can be attributed to the activation component of PANAS-PA, which is likely to be higher during weekdays than weekends.

Egloff et al. (1995) also found different effects of time of day on PANAS-PA and other measures of PA. Whereas PANAS-PA increases from morning until early afternoon and then decreases in the evening, other PA scales do not show the same pattern. Again, this finding can be explained by the inclusion of activation items in the PANAS. Items such as alert reflect energetic arousal (Schimmack & Reisenzein, 2002; Thayer, 1989), and it is well known that energetic arousal has a circadian rhythm. However, pleasant affects such as happy do not show the same rhythm. Thus, scales that focus on positive valence are unlikely to show a circadian rhythm.

Schimmack (2003) compared the predictive validity of various PA items for life satisfaction judgments. Affect was assessed in an experience sampling study. The aggregated ratings of momentary happiness were a better predictor of life satisfaction than a scale that combined happiness with pride and affection. Thus, the PANAS-PA scale, which intentionally does not include happiness or related items, is likely to underestimate the contribution of affective experiences to well-being.

PANAS-PA, but have no effect on PANAS-NA, and negative pictures influence PANAS-NA but have no effect on PANAS-PA. Contrary to this prediction, negative pictures produced an increase in PANAS-PA. Patrick and Lavoro (1997) found that the increase in PANAS-PA in response to negative pictures was due to activation items of the PANAS-PA scale such as alert, attentive, and interested. These items reflect engagement with a stimulus but do not reveal the evaluation of a stimulus. Patrick and Lavoro's (1997) findings suggest that the PA scale of the PANAS does not provide unambiguous information about SWB because it includes activation items that reflect engagement with a stimulus independent of valence.

In sum, different PA and NA measures are likely to produce similar results because they show high convergent validity. However, different PA and NA measures are not identical and do produce different results in some situations. Thus, the best strategy might be a multidimensional assessment of PA and NA and a careful examination of specific affects (Diener et al., 1995).

Validity of Different Methods

Besides item content, researchers have to choose between methods for the assessment of PA and NA. The most common methods are self-reports, in-

formant reports, and aggregated momentary/diary data. Each method has advantages and disadvantages. Thus, the best assessment of SWB is based on a multimethod assessment (Diener et al., 1995). However, multimethod assessments are costly and impractical in routine studies of SWB. Thus, it is important to examine the strengths and weaknesses of individual methods.

One approach examines the correlations of different methods to each other. A method with a larger amount of valid variance should be correlated more highly with other methods. Diener et al.'s (1995) multimethod study revealed that the correlation between self-reports and daily diary data of PA ($r = .68$) was higher than the correlations of these two methods with informant reports ($r = .53$, .55, respectively). For NA, correlations between self-reports and daily diary data ($r = .67$) were also higher than those with informant reports ($r = .39$, $r = .34$). This pattern of correlations suggests that retrospective self-reports and daily diary data contain equal amounts of valid variance, and more valid variance than informant reports. However, the comparison may be biased by systematic measurement variance that is common to retrospective self-reports and daily diary data. For example, socially desirable responding could inflate PA scores on both measures.

More conclusive evidence requires three truly independent methods. To my knowledge, the only study that fulfills this requirement is a multimethod study of personality traits that predict SWB, namely the depression facet of neuroticism and the positive emotion facet of extraversion (see Schimmack, Oishi, Furr, & Funder, 2004). Funder, Kolar, and Blackman [illegible]

tween mothers' and fathers' reports were $r = .65$ and $r = .52$ for depression and positive emotions, respectively.

Comparisons of retrospective and aggregated measures are virtually absent. Most studies have compared retrospective judgments to aggregated measures, assuming that aggregated daily ratings are more valid. However, as noted earlier, aggregated measures are biased by systematic measurement error and this bias may be more severe than the bias in retrospective ratings. Indeed, Watson and Tellegen (2002) suggested that aggregated measures that do not control for systematic measurement error are less valid than retrospective ratings. To examine this issue, I analyzed data from a daily diary study (see Schimmack, Oishi, Diener, & Suh, 2000 for details). In this study, participants rated the frequency of various emotions in the past 3 weeks. The first ratings were made before a 3-week daily diary study. Afterward, participants recorded the frequency of the same emotions each day for 3 weeks. At the end of the daily diary study, participants repeated retrospective ratings of the past 3 weeks. Importantly, these judgments covered the same 3-week period as the daily diary study. Thus, the second set of retrospective judgments is influenced by the daily recording of emotions (see Schimmack, 2002). Informant reports were used as an unbiased validation criterion. Informants completed questionnaires about life satisfaction (SWLS; Diener et al., 1985), and a short questionnaire of the Big Five. Extraversion and neuroticism were used as validation criteria because these two traits are strong predictors of SWB (Diener & Lucas, 1999).

[illegible]

(1995) reported the correlations between self-reports and three sets of informant reports, namely parents, college friends, and hometown friends, for trait measures of depression and positive emotions. For depression, the self-informant correlations ($rs = .27$ to $.43$) were stronger than the correlations among informant reports ($rs = .16$ to $.26$). For positive emotions, the self-informant correlations ($rs = .26$, $.34$, $.36$) were also stronger than the correlations among informant reports ($rs = .12$ to $.33$). This finding also suggests that self-reports of SWB contain more valid variance than informant reports. Funder et al. (1995) also explained why this is the case. Different informants see individuals in different contexts. Support for this hypothesis stems from much higher informant-informant correlations within each set of informants. For example, the correlations be-

Table 7.3 shows the results. All three measures of PA and NA predict informant reports of life satisfaction. All three measures of PA predict informant reports of extraversion, and all three measures of NA predict informant reports of neuroticism. This pattern is consistent with the theoretical model that extraversion is a disposition to experience more PA, neuroticism is a disposition to experience more NA, and life satisfaction is in part determined by the hedonic balance of PA and NA (Costa & McCrae, 1980; Diener & Lucas, 1999; Schimmack, Diener, et al., 2002). Hierarchical regression analyses revealed that retrospective postdiary ratings explained significantly more variance in neuroticism and extraversion than prediary reports. Daily diary ratings explained significantly more variance in life satisfaction and neuroticism scores than prediary ratings. Prediary self-ratings did not

TABLE 7.3 Prediction of Informant Reports of Life Satisfaction (SWLS), Extraversion, and Neuroticism

Criterion	Beta PA	Beta NA	R2
SWLS			
Self-report prediary	.19*	−.19*	.10
Self-report postdiary	.24*	−.20*	.12
Daily diary	.32*	−.34*	.15
Neuroticism			
Self-report prediary	−.08	.34*	.14
Self-report postdiary	−.18*	.35*	.19
Daily diary	−.28*	.43*	.18
Extraversion			
Self-report prediary	.23*	.01	.05
Self-report postdiary	.31*	−.07	.11
Daily diary	.19*	−.15	.04

*$p < .05$.

explain unique variance in criterion variables after controlling for postdiary ratings or daily diary data. Thus, the results suggest that daily diary data contain slightly more valid variance than retrospective ratings that are made without prior daily recording of emotions. However, even retrospective judgments prior to the daily diary study revealed predictive validity and are useful in routine studies of SWB.

Conclusion

This review examined methodological issues in the assessment of the affective component of SWB. Key conclusions were that PA and NA are separable components of SWB. The empirical relation between these dimensions depends on several methodological factors. Researchers should use asymmetric scales for the assessment of PA and NA, and they should be aware that aggregates of repeated assessments are likely to be biased by systematic measurement error. This systematic bias can be controlled in various ways (Schimmack, 2003; Watson & Tellegen, 2002). Researchers should also pay attention to the item content of PA and NA scales, as different scales sometimes produce different results. Multimethod studies are desirable, but studies with a single method are likely to produce valid results most of the time. One-time self-reports of SWB will remain the most commonly used method as they provide valid information and are much easier to obtain than other data.

References

Bagozzi, R. P., Wong, N., & Yi, Y. (1999). The role of culture and gender in the relationship between positive and negative affect. *Cognition and Emotion, 13*, 641–672.

Barrett, L. F., & Russell, J.A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*(4), 967–984.

Beebe-Center, J. G. (1932). *Psychology of pleasantness and unpleasantness*. New York: Van Nostrand.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401–423.

Clark, L. A., & Watson, D. (1988). Mood and the mundane: Relations between daily life events and self-reported mood. *Journal of Personality and Social Psychology, 54*, 296–308.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology, 38*, 668–678.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542–575.

Diener, E. (1999). Introduction to the special section on the structure of emotion. *Journal of Personality and Social Psychology, 76*, 803–804.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55,* 34–43.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71–75.

Diener, E., & Iran-Nejad, A. (1986). The relationship in experience between different types of affect. *Journal of Personality and Social Psychology, 50,* 1031–1038.

Diener, E., & Lucas, R. E. (1999). Personality and subjective well-being. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 213–229) New York: Russell Sage.

Diener, E., Smith, H., & Fujita, F. (1995). The personality structure of affect. *Journal of Personality and Social Psychology, 69,* 130–141.

Egloff, B., Tausch, A., Kohlmann, C., & Krohne, H. W. (1995). Relationships between time of day, day of the week, and positive mood: Exploring the role of the mood measure. *Motivation and Emotion, 19,* 99–110.

Eid, M., Notz, P., Schwenkmezger, P., & Steyer, R. (1994). Are mood dimensions monopolar? A review of empirical results and investigations with factor analyses of continuous and categorical variables as well as findings from a new study. *Zeitschrift fuer Differentielle und Diagnostische Psychologie, 15,* 211–233.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69,* 656–672.

Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology, 64,* 1029–1041.

Hemenover, S. H., & Schimmack, U. (2004). *That's disgusting! . . . but very amusing: Mixed feelings of amusement and disgust.* Manuscript under review.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2006). *Mixed moods: Affective responses to music with conflicting cues.* Manuscript under review.

disappointing wins and relieving losses. *Psychological Science, 15,* 325–330.

Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology, 71,* 616–628.

Meddis, R. (1972). Bipolar factors in affect adjective checklists. *British Journal of Social and Clinical Psychology, 11,* 178–184.

Patrick, C. J., & Lavoro, S. A. (1997). Ratings of emotional response to pictorial stimuli: Positive and negative affect dimensions. *Motivation and Emotion, 21,* 297–322.

Russell, J. A., & Carroll, J. M. (1999). The phoenix of bipolarity: Reply to Watson and Tellegen (1999). *Psychological Bulletin, 125,* 611–617.

Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings? Are semantic opposites mutually exclusive? *Cognition and Emotion, 15,* 81–97.

Schimmack, U. (2002). Frequency judgments of emotions: The cognitive basis of personality assessment. In P. Sedelmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 189–204). Oxford: Oxford University Press.

Schimmack, U. (2003). Affect measurement in experience sampling research. Journal of *Happiness Studies, 4,* 79–106.

Schimmack, U. (2004). Unpublished data.

Schimmack, U. (2005). Response latencies of pleasure and displeasure ratings: Further evidence for mixed feelings. *Cognition and Emotion, 19,* 671–691.

Schimmack, U., Bockenholt, U., & Reisenzein, R. (2002). Response styles in affect ratings: Making a mountain out of a molehill. *Journal of Personality Assessment, 78,* 461–483.

Schimmack, U., & Colcombe, S. (in press). *Eliciting mixed feelings with the paired-picture paradigm: A tribute to Kellogg (1915).* Cognition and Emotion.

Schimmack, U., Diener, E., & Oishi, S. (2002). Life-satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. *Journal of Personality, 70,* 345–385.

Schimmack, U., Oishi, S., & Diener, E. (2002).

Kellogg, C. E. (1915). Alternation and interference of feelings. *Psychological Monographs, 18,* 1–94.

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81,* 684–696.

Larsen, J. T., McGraw, A. P., Mellers, B. A., & Cacioppo, J. T. (2004). The agony of victory and thrill of defeat: Mixed emotional reactions to

Cultural influences on the relation between pleasant emotions and unpleasant emotions: Asian dialectic philosophies or individualism-collectivism? *Cognition and Emotion, 16,* 705–719.

Schimmack, U., Oishi, S., Diener, E., & Suh, M. (2000). Facets of affective experiences: A new look at the relation between pleasant and unpleasant affect. *Personality and Social Psychology Bulletin, 26,* 655–668.

Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and life satisfaction: A facet level analysis. *Personality and Social Psychology Bulletin, 30,* 1062–1075.

Schimmack, U., & Reisenzein, R. (2002). Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion, 2,* 412–417.

Stone, A. A., Hedges, S. M., Neale, J. M., & Satin, M. S. (1985). Prospective and cross-sectional mood reports offer no evidence of a "blue Monday" phenomenon. *Journal of Personality and Social Psychology, 49,* 129–134.

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

Thayer, R. E. (1989). *The biopsychology of mood and arousal.* London: Oxford University Press.

Watson, D. (1988). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of Personality and Social Psychology, 55,* 128–141.

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., & Tellegen, A. (1999). Issues in dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin, 125,* 601–610.

Watson, D., & Tellegen, A. (2002). Aggregation, acquiescence, and the assessment of trait affectivity. *Journal of Research in Personality, 36,* 589–597.

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality, 34,* 178–197.