Running head: LOW SELF-ESTEEM, STRESSFUL EVENTS, AND DEPRESSION

Disentangling the Effects of Low Self-Esteem and Stressful Events on Depression:

Findings from Three Longitudinal Studies

Ulrich Orth and Richard W. Robins

University of California, Davis

Laurenz L. Meier

University of Bern

Ulrich Orth
Department of Psychology
University of Bern
Muesmattstrasse 45
3012 Bern
Switzerland

E-mail: ulrich.orth@psy.unibe.ch
Phone: +41 31 631 40 43
Fax: +41 31 631 82 12

Abstract

Diathesis-stress models of depression suggest that low self-esteem and stressful events jointly influence the development of depressive affect. More specifically, the self-esteem buffering hypothesis states that, in the face of challenging life circumstances, individuals with low self-esteem are prone to depression because they lack sufficient coping resources whereas those with high self-esteem are able to cope effectively and consequently avoid spiraling downward into depression. The authors used data from 3 longitudinal studies of adolescents and young adults, who were assessed 4 times over a 3-year period (Study 1; $N = 359$), 3 times over a 6-week period (Study 2; $N = 249$), and 4 times over a 6-year period (Study 3; $N = 2,403$). In all three studies, low self-esteem and stressful events independently predicted subsequent depression, but did not interact in the prediction. Thus, the results did not support the self-esteem buffering hypothesis, but suggest that low self-esteem and stressful events operate as independent risk factors for depression. In addition, we found evidence in all three studies that depression, but not low self-esteem, is reciprocally related to stressful events, suggesting that individuals high in depression are more inclined to subsequently experience stressful events.

Key Words: self-esteem, depression, stressful events, diathesis-stress model

Disentangling the Effects of Low Self-Esteem and Stressful Events on Depression:

Findings from Three Longitudinal Studies

A growing body of research suggests that low self-esteem contributes to the development of depression. Overall, the findings support the "vulnerability model" which states that low self-esteem operates as a risk factor for depression (Beck, 1967; Metalsky, Joiner, Hardin, & Abramson, 1993). Importantly, many previous studies used prospective designs and controlled for prior levels of both self-esteem and depression (e.g., Kernis et al., 1998; Orth, Robins, & Roberts, 2008; Orth, Robins, Trzesniewski, Maes, & Schmitt, 2008; Roberts & Monroe, 1992). The effect of low self-esteem on depression holds for men and women, from adolescence to old age, and after controlling for content overlap between self-esteem and depression scales (Orth, Robins, & Roberts, 2008; Orth, Robins, Trzesniewski et al., 2008). Moreover, prior research has failed to support an alternative model of the relation between self-esteem and depression—the "scar model"—which hypothesizes that low self-esteem is an outcome rather than a cause of depression (Ormel, Oldehinkel, & Vollebergh, 2004; Orth, Robins, & Roberts, 2008; Orth, Robins, Trzesniewski et al., 2008).

The Diathesis-Stress Model

of Low Self-Esteem and Stressful Events

Many diathesis-stress models of depression consider low self-esteem to be a predisposing factor for the development of depression (e.g., Beck, 1967; Brown & Harris, 1978; Hammen, 2005; Metalsky et al., 1993). In the face of challenging life circumstances, individuals with low self-esteem are assumed to have fewer coping resources and thus are prone to depression whereas those with high self-esteem are assumed to have better coping resources and thus avoid spiraling downward into depression (see Hypothesis 1 in Figure 1). In other words, the

experience of stressful events generally contributes to depression, but individuals with relatively

high self-esteem are buffered against this effect (and, conversely, individuals with relatively low

self-esteem are more vulnerable to this effect). If self-esteem buffers individuals against the

deleterious consequences of stressful life events, then low self-esteem and stressful events should

have an interactive effect on subsequent depression.[1] The buffering hypothesis is a commonly

accepted view of the causal relationship among self-esteem, stressful events, and depression. For

example, Abela, Webb, Wagner, Ho, and Adams (2006) state that following negative events

"protective factors, such as high self-esteem, may prevent the outcome of depressive symptoms

by decreasing the negative impact of depressogenic thoughts on the affective, cognitive,

behavioral, and physiological symptoms of depression" (p. 329), and Roberts (2006), in a review

on self-esteem from a clinical perspective, states that one of the themes that emerges from the

depression literature is that self-esteem "interacts with other risk factors, such as life stress and

attributional style, in the prediction of depression" (p. 300).

Numerous studies have tested the self-esteem buffering hypothesis, that is, whether the

effects of stressful events on subsequent depression (controlling for prior levels of depression)

were stronger for low vs. high self-esteem individuals. Table 1 provides a summary of the

findings from these previous studies. Four studies confirmed the hypothesized interaction

(Abela, 2002; Fernandez, Mutran, & Reitzes, 1998; Metalsky et al., 1993; Ralph & Mineka,

1998); seven studies failed to find the hypothesized interaction (Butler, Hokanson, & Flynn,

1994; Cheng & Lam, 1997; Kernis et al., 1998; Lewinsohn, Hoberman, & Rosenbaum, 1988;

Murrell, Meeks, & Walker, 1991; Roberts & Monroe, 1992; Southall & Roberts, 2002); three

studies failed to find the hypothesized two-way interaction but found evidence of a three-way

interaction of self-esteem, stressful events, and third variables such as dysfunctional attitudes

(Abela & Skitch, 2007; Abela et al., 2006; Robinson, Garber, & Hilsman, 1995); and one study reported an interaction effect showing the opposite pattern (the effect of stressful events on depression was stronger for high vs. low self-esteem individuals, Whisman & Kwon, 1993).

Thus, previous research testing the interactive effects of self-esteem and stressful events has yielded highly inconsistent results. Table 1 shows that the studies used a wide range of samples (e.g., different age groups) and designs (e.g., the prospective time intervals ranged from a few days to 2 years). However, no consistent pattern emerges from these study characteristics that might help to explain the divergent pattern of results.

From a methodological perspective, it seems possible that the available evidence is biased by a file-drawer effect, such that studies that failed to support the buffering hypothesis were less likely to be published. Moreover, few published studies have sufficient power to detect interaction effects, because: (a) interactions are typically weak in magnitude (Chaplin, 1991) and (b) the interaction term in a multiple regression analysis is typically plagued by low reliability, reflecting the product of the reliability of the two variables and producing a downward bias in the interaction effect (Cohen, Cohen, West, & Aiken, 2003; Jaccard & Wan, 1995).

The present research, in contrast, uses latent variable modeling to separate construct variance from measurement error, leading to increased statistical power to test interactions and less biased estimates of the magnitude of the interaction effect (Jaccard & Wan, 1995). Moreover, we test the self-esteem buffering hypothesis in three independent longitudinal studies.

<p style="text-align:center">Additional Hypotheses Concerning the Relations</p>

<p style="text-align:center">Among Low Self-Esteem, Stressful Events, and Depression</p>

In addition to testing the self-esteem buffering hypothesis (Hypothesis 1), the present research tests several additional hypotheses. Hypothesis 2 (see Figure 1) specifies that stressful

events mediate the effect of low self-esteem on depression. Specifically, individuals with low self-esteem might generate stressful events through their own behavior, which in turn contributes to depression. For example, low self-esteem is prospectively linked to aggression and other antisocial behaviors that might contribute to interpersonal conflicts and other stressful life events (Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005). In close relationships, low self-esteem individuals tend to perceive their partner's behavior more negatively, thereby increasing the likelihood of relationship conflicts and rejection (Murray, Holmes, & Griffin, 2000; Murray, Rose, Bellavia, Holmes, & Kusche, 2002). We know of no previous research that has systematically tested whether low self-esteem contributes to stressful experiences. There is, however, considerable evidence that depressed individuals experience more stressful life events (Cole, Nolen-Hoeksema, Girgus, & Paul, 2006; Davila, Bradbury, Cohan, & Tochluk, 1997; Hammen, 1991; Holahan, Moos, Holahan, Brennan, & Schutte, 2005; Kim, Conger, Elder, & Lorenz, 2003). Because the link between low self-esteem and stress generation has not been tested, it is possible that depressed individuals experience more stressful events at least in part because they tend to have low self-esteem. To address this possibility, the present research tests whether low self-esteem and depression have independent effects on the occurrence of stressful events. If low self-esteem contributes to stressful experiences, we can then test whether stressful events mediate the effect of low self-esteem on depression.

Alternatively, reversing the causal chain, low self-esteem might mediate the effects of stressful events on depression (see Hypothesis 3 in Figure 1). That is, certain stressful events (e.g., relationship break-up, academic failures) may lead to low self-esteem which in turn contributes to depressive affect (Oatley & Bolton, 1985; Roberts & Monroe, 1999). Although we know of no previous prospective study that has directly tested this possibility, Tram and Cole

(2000) found that self-perceived competence (i.e., a construct related to self-esteem) partially

mediated the effect of stressful life events on subsequent depression in a sample of adolescents.

Hypotheses 4 and 5 concern possible spurious effects (see Figure 1). Hypothesis 4

specifies that stressful events lead to both low self-esteem and higher levels of depression,

creating a spurious relation between the two. To test this possibility, we examine whether the

relation between self-esteem and depression holds after statistically controlling for stressful

events. Hypothesis 5 specifies that low self-esteem leads to stressful events and depression,

creating a spurious relation between the two. The present research will test whether the relation

between stressful events and depression holds after controlling for the effect of self-esteem.

<div align="center">The Present Research</div>

All hypotheses described above can be tested in one overarching longitudinal model,

which analyzes prospective main, interactive, and mediation effects. Figure 2 provides a generic

illustration of this model. The self-esteem buffering hypothesis can be tested as interaction

between self-esteem at one time point (e.g., Time 1) and stressful events in the subsequent time

interval (e.g., Time 1 to 2) predicting depression at the subsequent time point (e.g., Time 2),

controlling for prior depression (e.g., Time 1). The mediation hypotheses (i.e., stressful events as

a mediator of the effect of low self-esteem on depression; low self-esteem as a mediator of the

effect of stressful events on depression) can be tested following the guidelines outlined by Cole

and Maxwell (2003). The spurious effects hypotheses (i.e., stressful events confound the relation

between self-esteem and depression; self-esteem confounds the relation between stressful events

and depression) can be tested by analyzing whether effects in the trivariate model differ from

effects in bivariate models of the constructs. The model shown in Figure 2, and therefore all of

the hypotheses, can be tested in all three longitudinal studies examined in the present research.

This research extends previous studies on self-esteem, stressful events, and depression in several ways. First, in contrast to previous studies, we investigated reciprocal effects between all three constructs. By doing so, we were able to simultaneously test alternative models. Second, in contrast to most previous studies, we used more appropriate statistical methods based on latent variable modeling, providing better estimates of the effects and more flexibility in controlling for antecedent and concurrent effects (Cole & Maxwell, 2003; Finkel, 1995). Third, in Study 1 and Study 3 we used data sets that include multiple repeated assessments (i.e., four waves), so the main and interaction effects could be replicated and constrained across multiple time points, increasing both the reliability and validity of the estimates; most previous studies used two-wave longitudinal designs, or, if more than two waves were available, most previous studies did not take full advantage of the design by testing whether effects may be constrained across time. Fourth, in Study 2 stressful events were assessed at measurement occasions that were non-overlapping with measurement occasions for self-esteem and depression; consequently, any occasion effects (e.g., mood) that might have influenced the assessments of self-esteem and depression would be independent of occasion effects influencing reports of stressful events. Fifth, we cross-validated our results using three data sets with different sample and design characteristics; by replicating the findings across studies, we reduce methodological concerns unique to each study and strengthen confidence in the overall pattern of results.

We decided to focus on adolescence and young adulthood for two reasons. First, the prevalence of depressive disorders is high during adolescence and young adulthood (Costello, Erkanli, & Angold, 2006; Kessler et al., 2005), so this developmental period is particularly important for understanding the underlying etiology of depression. Second, self-esteem and depression are particularly likely to show changes during adolescence and young adulthood

because of the many transitions that occur during this time of life (Galambos, Barker, & Krahn, 2006; Mirowsky & Kim, 2007; Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002). The adolescent period is associated with rapid maturational changes, shifting societal demands, conflicting role demands, and increasingly complex romantic and peer relationships. After graduating from high school, many young adults move away from home for the first time, begin college and full-time jobs, or marry and have children. The developmental process of becoming an adult often entails a questioning of one's identity and subsequent reformulation of conceptions and evaluations of the self. Thus, the developmental period of adolescence and young adulthood is ideally suited to test hypotheses concerning the relations among low self-esteem, stressful events, and depression.

Study 1

Study 1 used data from the Berkeley Longitudinal Study, an ongoing study of a cohort of individuals who entered the University of California at Berkeley in 1992 (for further information see Robins, Hendin, & Trzesniewski, 2001; Robins, Noftle, Trzesniewski, & Roberts, 2005).[2] Six assessments were conducted over a 4-year period: first week of college, end of first semester, and end of first, second, third, and fourth years of college. We focused our analyses on the latter four assessments (denoted as Time 1 to Time 4 in the remainder of this article) because stressful events and depression were not assessed in the first two assessments. To reduce age heterogeneity, we restricted the sample to participants who were 18- or 19-years-old when they entered college.

Method

Participants

The sample consisted of 359 individuals (59% female). Mean age of participants at Time 1 was 18.3 years ($SD = 0.5$, Range = 18 to 19). Forty-three percent were Asian, 31% Caucasian, 13% Chicano/Latino, 5% African American, 1% American Indian, 2% of other ethnicity, and 5% did not specify ethnicity.

Data were available for 270 individuals at Time 1, 232 individuals at Time 2, 177 individuals at Time 3, and 277 individuals at Time 4. To investigate the potential impact of attrition, we tested for differences on study variables between participants who completed the Time 4 assessment and participants who dropped out of the study before Time 4. For one variable (i.e., stressful events reported at Time 2), participants who dropped out reported higher values than participants who completed the full study ($d = 0.52$, $p < .05$). No significant differences emerged for any of the other variables (i.e., self-esteem at Times 1 to 3, depression at Times 1 to 3, and stressful events at Time 3).

Measures

Self-esteem. Self-esteem was assessed with the 10-item Rosenberg Self-Esteem Scale (RSE, Rosenberg, 1965). The RSE is the most commonly used and well-validated measure of global self-esteem (Robins et al., 2001). Responses were measured on a 5-point scale ranging from 1 (*not very true of me*) to 5 (*very true of me*). The alpha reliability of the RSE was .89 at Time 1, .91 at Time 2, .90 at Time 3, and .90 at Time 4.

Stressful events. At Times 2 to 4, the occurrence of 12 stressful events during the past year was assessed. The items were: "breaking up with boyfriend/girlfriend," "death or serious illness/injury of a close family member or friend," "serious personal illness or injury," "fired from job or serious trouble with employer," "transferred out of UC Berkeley to a different college," "academic probation," "victim of a crime," "failing an important exam," "failing a

course," "financial problems concerning school (e.g., in danger of not having sufficient money to continue)," "arrested for a crime," and "dropped out of college". For the analyses, we used the summed number of event, with a possible range from 0 to 12. [Coefficient alpha, which was .38 at Time 2, .30 at Time 3, and .49 at Time 4, is not an appropriate indicator of reliability because the heterogeneous items do not measure an internally consistent construct, but rather function as an index of the degree of cumulative life stress (see Bollen & Lennox, 1991; Streiner, 2003).]

Depression. Depression was assessed with the Center for Epidemiologic Studies Depression Scale (CES-D, Radloff, 1977). The CES-D is a frequently used 20-item self-report measure for the assessment of depressive symptoms in non-clinical, sub-clinical, and clinical populations, and its validity has been repeatedly confirmed (Eaton, Smith, Ybarra, Muntaner, & Tien, 2004). Participants were instructed to assess the frequency of their reactions within the preceding seven days. Responses were measured on a 4-point scale (0 = *rarely or none of the time, less than one day*, 1 = *some or a little of the time, one to two days*, 2 = *occasionally or a moderate amount of time, three to four days*, 3 = *most or all of the time, five to seven days*). The alpha reliability of the CES-D was .91 at Time 1, .91 at Time 2, .90 at Time 3, and .91 at Time 4.

Procedure for the Statistical Analyses

The analyses were conducted using the Mplus 5.2 program (Muthén & Muthén, 2007). To deal with missing values, we employed maximum likelihood estimation, which produces less biased and more reliable results compared to conventional methods of dealing with missing data, such as listwise or pairwise deletion (Schafer & Graham, 2002). Models including latent interactions were estimated by numerical integration using the Monte Carlo algorithm with 2,000 integration points.[3]

Model fit was assessed by the Tucker-Lewis-Index (TLI), the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA), based on the recommendations of Hu and Bentler (1999) and MacCallum and Austin (2000). Hu and Bentler (1999) suggest that good fit is indicated by values greater than or equal to .95 for TLI and CFI, and less than or equal to .06 for RMSEA. Because these indices are not available for models including latent interactions, we also report the Bayesian Information Criterion (BIC). For BIC, absolute values cannot be interpreted, but when comparing models, lower values indicate better model fit.

Results and Discussion

Table 2 shows means and standard deviations of the measures used in Study 1. For the latent variables self-esteem and depression, we used item parcels as indicators because they produce more reliable latent variables than individual items (Little, Cunningham, Shahar, & Widaman, 2002). For both self-esteem and depression, we randomly aggregated the items into three parcels. In contrast, for stressful events, parceling was not possible because, as discussed above, stressful event checklists do not measure an internally consistent construct. We therefore used the stressful event scale as a single indicator of the latent variable, and modeled the measurement error of the indicator by fixing the residual variance to 20% of the total variance of the stressful event scale (corresponding to an assumed reliability of the scale of .80).[4]

Before testing the study hypotheses, we first tested a basic model that did not include the interaction between self-esteem and stressful events. For self-esteem and depression, the uniquenesses of individual indicators were correlated across time to control for bias due to indicator-specific variance (cf. Cole & Maxwell, 2003). (This procedure was not appropriate for stressful events, which were measured by single indicators.) For each construct, the loading of

the first indicator was set to 1 to identify the model. For self-esteem and depression, the factor

loadings of the second and third indicator were constrained to be equal across time to ensure that

the latent constructs were measured similarly across occasions. We accounted for variance due to

specific measurement occasions by cross-sectionally correlating the disturbances of self-esteem

and depression (cf. Cole & Maxwell, 2003). However, no cross-sectional correlations were

included for stressful events and self-esteem or for stressful events and depression because the

model included direct effects between the constructs, which supersede correlated disturbances.

A model constraining the stability and cross-lagged effects to be equal across the three

time intervals did not lead to a significant reduction in fit relative to the unconstrained model,

$\Delta\chi^2(17, N = 359) = 9.8$, *ns*. Consequently, we included these equality constraints in the

remainder of the analyses. Table 3 shows that the overall fit of the no interaction model was

good. Table 4 shows the unstandardized estimates and standard errors of the structural

coefficients.

We tested two models of the interactive effect of self-esteem and stressful events on

depression, one that constrained the effect to be equal across assessments and one that did not. In

the constrained model, the interaction effect was close to zero and nonsignificant. In the

unconstrained model, one of the three interactions was significant (Table 4), but the other two

interaction effects were in the opposite direction of the predicted effect. Moreover, the

unconstrained model fit the data worse than the constrained model (as indicated by a higher

value for BIC; see Table 3). Two other results shown in Table 4 are noteworthy: First, the main

effects were virtually identical for the models with and without an interaction effect. Second, the

standard errors were small, for both main effects and interaction effects, indicating successful

model estimation.

Thus, the results suggest that self-esteem and stressful events did not interact in predicting subsequent depression. Moreover, the no interaction model fit the data better than the interaction models (as indicated by a lower value for BIC; see Table 3). We therefore examined the standardized estimates for the no interaction model (see Figure 3). The results showed that the stability effects were significant for all three constructs, and that three cross-lagged effects were significant: both self-esteem and stressful events predicted subsequent depression, and depression predicted subsequent stressful events.[5] We also tested for gender differences in the structural coefficients, using a multiple group analysis. A model allowing for different coefficients for male and female participants did not significantly improve model fit, relative to a model with constraints across gender; the difference in model fit was $\Delta\chi^2(13, N = 359) = 21.5$, *ns*.[6]

In addition to the self-esteem buffering hypothesis, we tested several additional hypotheses. First, the effect of self-esteem on depression was not mediated by stressful events (Hypothesis 2), as indicated by a nonsignificant Sobel Test. Thus, the results suggest that low self-esteem did not have a stress-generating effect. For comparison purposes, we inspected the effect sizes in a bivariate model of self-esteem and stressful events, omitting depression. The effect of self-esteem on stressful events was small (at -.05, -.04, and -.04 for the three time intervals, respectively; all coefficients *ns*). Thus, even when the relations between self-esteem and stressful events were analyzed separately from depression, low self-esteem did not predict stressful events.

Second, the effect of stressful events on depression was not mediated by low self-esteem (Hypothesis 3), as indicated by a nonsignificant Sobel Test. Again, for comparison purposes, we inspected the effect sizes in a bivariate model of self-esteem and stressful events, omitting

depression. The effect of stressful events on self-esteem was very small and virtually at the same size as in the trivariate model (at -.04 for all three time intervals; all coefficients *ns*).[7]

Third, the effect of self-esteem on depression held after controlling for stressful events (Hypothesis 4). In fact, the effect of self-esteem on depression was as strong as in a bivariate model of self-esteem and depression (-.20, -.21, and -.22 for the three time intervals, respectively, all ps < .05). Thus, the findings further support the influence of low self-esteem on depression, because they rule out one plausible third-variable confound—stressful life events.

Fourth, the results replicate the effect of depression on stress generation (see Cole et al., 2006; Grant, Compas, Thurm, McMahon, & Gipson, 2004; Hammen, 1991), and show that this effect remains significant after controlling for self-esteem (Hypothesis 5). Thus, the findings suggest that depression and stressful events have reciprocal prospective relations—individuals experiencing stressful events are more likely to increase in depressive affect and individuals high in depressive affect are more likely to experience stressful events. Interestingly, controlling for self-esteem actually increased the size of these reciprocal effects, suggesting that self-esteem serves as a suppressor variable in this context (Paulhus, Robins, Trzesniewski, & Tracy, 2004); when self-esteem was removed from the model, the effect of stressful events on depression (at .16, .15, and .18 for the three time intervals, respectively; all *p*s < .05) and the effect of depression on stressful events (at .16, .14, and .15 for the three time intervals, respectively; all *p*s < .05) were slightly smaller.

To cross-validate the findings of Study 1, we conducted two additional longitudinal studies. Study 2 allowed us to address a methodological limitation of Study 1 related to the timing of the assessment of stressful events: although Study 1 participants were asked to report

life events that occurred during the past year, they nonetheless reported those events and responded to the self-esteem and depression questionnaires in the same assessment.

Study 2

In Study 2, self-esteem and depression were assessed on two occasions separated by six weeks (denoted as Time 1 and Time 3). In contrast to Study 1, stressful events were assessed repeatedly during the intervening period (denoted as Time 2), rather than at the same time as the assessments of self-esteem and depression. This procedure has several advantages. First, any occasion effects (e.g., mood) that might have influenced the assessments of self-esteem and depression would be independent of occasion effects influencing reports of stressful events. Second, by using multiple assessments (up to 12 in the present study) of stressful events, the reliability of the variable was significantly increased. Third, the procedure allowed us to assess a different type of stressful events: whereas Study 1 examined stressful life events (some of which were major life events, such as criminal victimization or death of a close family member), Study 2 examined daily hassles at the workplace. Some researchers have argued that daily hassles actually have a stronger impact on psychological adjustment than major life stressors (Lazarus, 1999).

Method

Participants and Procedures

The sample consisted of 249 trainees (36% female) at a large Swiss company. Mean age of participants at Time 1 was 18.0 years ($SD = 1.3$, Range = 16 to 23). The data were collected using web-based questionnaires which were accessible only to individuals who were invited to participate (individualized links were emailed to 270 trainees). Participants were assured that

their data would be kept completely confidential and would not be shared with their supervisors or anyone else.

Data were available for 221 individuals at Time 1, 197 individuals at Time 2, and 185 individuals at Time 3. To investigate the potential impact of attrition, we tested for differences on study variables between participants who completed the Time 3 assessment and participants who dropped out of the study before Time 3. No significant differences emerged for any variables.

Measures

Self-esteem. Self-esteem was assessed with a German version of the 10-item RSE (von Collani & Herzberg, 2003). Responses were measured on a 6-point scale ranging from 0 (*not very true of me*) to 5 (*very true of me*). The alpha reliability was .86 at Time 1 and .89 at Time 3.

Stressful events. During the first 12 workdays after the Time 1 assessment, participants were asked to report the occurrence of stressful events at the workplace. Assessments were conducted at 11.30 a.m., and participants reported events that occurred earlier in the morning. The five events were: "I was criticized for my work," "I made a mistake that will have consequences," "I was left alone in a difficult situation," "I had a conflict/argument with another person," and "I was unfairly treated". Because most of the trainees had to attend school on some of the weekdays, we expected 6 daily reports for each participant. However, for practical reasons, participants received emails providing access to the questionnaire on every weekday; therefore, the maximum number of daily reports was 12. To increase reliability, the stressful events scale was computed only if four or more daily reports were available ($N = 156$). The average number of daily reports was $M = 6.1$ ($SD = 1.4$, Range 4 to 12). We computed the stressful events scale separately for each day, with a possible range from 0 to 5 events

experienced that day. Coefficient alpha for the 5 events was .61; however, as noted in Study 1, coefficient alpha is not an appropriate indicator of reliability for event scales.

Depression. Depression was assessed with the German 15-item short form of the CES-D (Hautzinger & Bailer, 1993). Participants were instructed to assess the frequency of their reactions within the preceding seven days. Responses were measured on a 4-point scale (0 = *rarely or none of the time*, 1 = *sometimes*, 2 = *frequently*, 3 = *most of the time*). The alpha reliability was .92 at both Time 1 and 3.

Procedure for the Statistical Analysis

All analyses were conducted using Mplus 5.2. Model fit was assessed using the same procedures as in Study 1.

Results and Discussion

Table 5 shows means and standard deviations of the measures used in Study 2. The models tested were identical to Study 1, except that the Study 2 models included only two instead of four waves of data (so that only one interaction model could be tested).

Overall, the results replicated the findings of Study 1 (see Table 6 and Figure 4). First, we again failed to find evidence that self-esteem buffers the effects of stressful events on depression (Hypothesis 1); model fit was worse with an interaction term than without one (as indicated by a lower BIC value; see Table 3) and the interaction effect was nonsignificant. Thus, stressful events contribute to depression, regardless of whether the individual is high or low in self-esteem. Second, the effect of self-esteem on depression was not mediated by stressful events (Hypothesis 2), as indicated by a nonsignificant Sobel Test. As in Study 1, the findings support the stress-generation hypothesis for depression but not for low self-esteem; moreover, we also replicated the finding from Study 1 that the effect of depression on stressful events is even

stronger after controlling for self-esteem (.33 vs. .18). Third, the effect of stressful events on

depression was not mediated by low self-esteem (Hypothesis 3). Fourth, the effect of self-esteem

on depression remained significant after controlling for stressful events (Hypothesis 4). Fifth, the

effect of depression on stressful events remained significant after controlling for self-esteem

(Hypothesis 5). Finally, as in Study 1, a model allowing for different coefficients for male and

female participants did not significantly improve model fit relative to a model with constraints

across gender, $\Delta\chi^2(10, N = 249) = 15.8$, ns.[8]

Thus, Studies 1 and 2 both failed to support the interactive effect of self-esteem and

stressful events on depression. However, it is crucial to evaluate whether we had sufficient

statistical power to detect interactions, to avoid falsely accepting the null hypothesis that no

interaction exists. Under the assumption that predictors are measured without error—an

assumption that is acceptable in latent variable modeling because construct factors and

measurement error are explicitly modeled—a sample size of at least 26 is required for large

interaction effects, 55 for medium interaction effects, and 392 for small interaction effects

(Cohen et al., 2003). Thus, the sample sizes of the Studies 1 and 2, which were 359 and 249,

respectively, were sufficiently large to detect medium but not small interaction effects. However,

in psychological research interaction effects are generally small in magnitude (Chaplin, 1991;

Cohen et al., 2003). Therefore, to address potential concerns related to statistical power, Study 3

analyzed data from a very large longitudinal study.

Study 3

Study 3 used data from the National Longitudinal Survey of Youth (NLSY79), a national

probability survey which was started in 1979 (for further information about this study, see Center

for Human Resource Research, 2006).[9] The present analyses focused on the children of study

participants, who were first assessed in 1994 if they had reached the age of 15 years. These

adolescents and young adults were assessed biennially from 1994 to 2004, resulting in six

assessments. However, the number of assessments available for each participant varies widely

because there is a complex pattern of planned missing data due to budgetary reasons. For

example, in 1998, only children aged 15 to 20 were interviewed and, in 2000, about 40% of the

black and Hispanic oversamples were not surveyed. Moreover, because at every assessment

additional children reached the age of 15 years and thus became eligible for assessment, the

sample size increased with every assessment (*N*s ranged from 980 in 1994 to 5024 in 2004). The

design of the study produced substantial age heterogeneity (e.g., participants in the 2004

assessment ranged in age from 15 to 34 years). To reduce the age heterogeneity of the sample,

we decided to analyze sequences of four repeated assessments for those individuals who began

the survey in 1994, 1996, or 1998 at the age of 15 or 16. The data for these three cohorts were

restructured so that the age of every individual was 15 or 16 at Time 1.

Method

Participants

The sample consisted of 2,403 individuals (50% female). Mean age of participants at the

first assessment was 15.5 years (*SD* = 0.5, Range = 15 to 16). Sixty-one percent were White

(non-Hispanic), 21% were Black, 12% were Hispanic, 2% were American Indian, and 4% were

of other ethnicity. Data on study variables were available for 2,094 individuals at Time 1, 1,923

individuals at Time 2, 1,894 individuals at Time 3, and 2,151 individuals at Time 4. To

investigate the potential impact of attrition we tested for differences on study variables between

participants who completed the Time 4 assessment and participants who dropped out of the study

before Time 4. For one variable (i.e., depression reported at Time 1), participants who dropped

out reported lower values than participants who completed the full study ($d = $ -0.15, $p < .05$). No

significant differences emerged for any of the other variables (i.e., self-esteem at Times 1 to 3,

depression at Times 2 to 3, and stressful events at Times 2 to 3).

Measures

Self-esteem. As in Studies 1 and 2, self-esteem was assessed with the 10-item RSE.

Responses were measured on a 4-point scale ranging from 1 (*strongly disagree*) to 4 (*strongly

agree*). The alpha reliability was .84 at Time 1, .87 at Time 2, .88 at Time 3, and .88 at Time 4.

Stressful events. At Times 2 to 4, the occurrence of five stressful events during the past

two years was assessed. The items were: "dropped out of regular school for at least one month

and then returned," "repeated a grade in school," "had any accidents or injuries that required

medical attention," "had any illnesses that required medical attention or treatment," and "been

convicted of any charges other than a minor traffic violation".[10] For the analyses, we used the

summed number of events, with a possible range from 0 to 5. Coefficient alpha was .39 at Time

2, .14 at Time 3, and .09 at Time 4; however, as noted in Study 1, coefficient alpha is not an

appropriate indicator of reliability for event scales.[11]

Depression. As in Studies 1 and 2, depression was assessed with the CES-D. The

NLSY79 uses a short version of the CES-D with 7 items: "I did not feel like eating; my appetite

was poor", "I had trouble keeping my mind on what I was doing", "I felt depressed", "I felt that

everything I did was an effort", "My sleep was restless", "I felt sad", "I could not get 'going'".

For each item, participants were instructed to assess the frequency of their reactions within the

preceding seven days. Responses were measured on a 4-point scale (0 = *rarely, none of the time,

one day*; 1 = *some, a little of the time, one to two days*; 2 = *occasionally, moderate amount of the

time, three to four days*; 3 = *most, all of the time, five to seven days*). The alpha reliability of this

short form of the CES-D was .65 at Time 1, .66 at Time 2, .67 at Time 3, and .68 at Time 4. [In

Study 1, which used the full 20-item version of the CES-D, the 7-item version correlated .92 at

Time 1, .91 at Time 2, .90 at Time 3, and .92 at Time 4 with the full scale.]

Procedure for the Statistical Analyses

All analyses were conducted using Mplus 5.2. Model fit was assessed using the same

procedures as in Study 1.

Results and Discussion

Table 7 shows means and standard deviations of the measures used in Study 3. The

models tested were identical to Studies 1 and 2. As in Study 1, we tested whether the stability

and cross-lagged effects could be set equal across the three time intervals without a significant

decrease in model fit. The difference in model fit of the constrained and unconstrained model

was significant with $\Delta\chi^2(17, N = 2,403) = 70.1, p < .05$. However, with sufficiently large

samples, the chi-square difference test for nested models will always be significant, even when

the true difference in fit is very small and theoretically irrelevant (MacCallum, Browne, & Cai,

2006). The values for RMSEA, which is less sensitive to sample size (cf. MacCallum et al.,

2006), differed by only .001 between the unconstrained and constrained model. Consequently,

we used constraints on all repeated structural coefficients in the remainder of the analyses.

The results closely replicated the findings of Studies 1 and 2 (see Table 8 and Figure 5).

Most importantly, we again failed to find evidence that self-esteem buffers the effects of stressful

events on depression; model fit was worse for models with interactions than without (as

indicated by higher BIC values; see Table 3) and the interaction effects were nonsignificant.

Likewise, the pattern of results was similar to Studies 1 and 2 with regard to the mediation

hypotheses (Hypotheses 2 and 3) and the confounding hypotheses (Hypotheses 4 and 5). The

only difference between the results of Study 3 and Studies 1 and 2 was that the effects of stressful events on self-esteem were significant—however, these effects were very small (at -.03 to -.04) and their statistical significance is due to the very large sample size of Study 3.

Finally, as in Studies 1 and 2, we tested a model allowing for different coefficients for male and female participants. The chi-square difference was significant with $\Delta\chi^2(13, N = 2,403) = 31.2, p < .05$; however, because of the very large sample size, we again inspected values for RMSEA, which were identical for both the unconstrained and constrained model. Consequently, we concluded that male and female participants did not differ meaningfully in the structural coefficients.

General Discussion

In the present research, we investigated the main and interactive effects of low self-esteem and stressful events on depression, using three longitudinal studies of adolescents and young adults. The results from these studies provide converging support, or lack of support, for several hypotheses concerning the relations among the three constructs.

The results did not support the self-esteem buffering hypothesis, which states that the effects of stressful events (i.e., stressful life events or daily hassles) on subsequent depression are stronger for low vs. high self-esteem individuals (Hammen, 2005; Metalsky et al., 1993). As reviewed in the introduction, previous empirical tests of this hypothesis have yielded inconsistent results, which may have been due to the fact that the traditional strategy of testing interaction effects (i.e., computing the product of manifest variables) suffers from unreliable coefficients and low statistical power. The present research bypassed these methodological problems by using a latent variable approach, and using sample sizes that provided sufficient statistical power to detect small (Study 1), medium to small (Study 2), and very small (Study 3) interaction

effects. In Studies 1 and 3, statistical power was further enhanced by constraining the interaction effect across three time intervals (which reduces the standard errors of the estimates and increases the power of significance tests). Finally, the combined statistical power to detect an interaction effect in at least one out of three studies was even larger compared to the power of each individual study; however, in all three studies the interaction effects of self-esteem and stressful events were nonsignificant. Thus, the present research provides compelling evidence that, contrary to the diathesis-stress model, low self-esteem is a risk factor for depression whose strength is not moderated by the presence or absence of stressful events, and, similarly, that stressful events are a risk factor for depression whose strength does not depend on level of self-esteem. Although we failed to find support for the self-esteem buffering hypothesis, we believe that these results provide a significant contribution to the field, given the widespread belief in the validity of the hypothesis; many researchers have argued that it is critical for scientific fields to publish null results, when there is a clear rationale for the hypothesis and the research is well-conducted and has sufficient statistical power (Cooper, DeNeve, & Charlton, 1997; Fraley & Marks, 2007; Greenwald, 1975).

The results of the three longitudinal studies also failed to confirm the hypothesis that stressful events mediate the effects of low self-esteem on depression, or that low self-esteem mediates the effects of stressful events on depression. Therefore, future research should test other mediating factors. For example, regarding the effects of low self-esteem on depression, such factors include both intrapersonal processes (e.g., low self-esteem likely elicits rumination about negative aspects of the self, which in turn increases and prolongs negative affect, see Nolen-Hoeksema, 2000) and interpersonal processes (e.g., low self-esteem motivates social avoidance,

thereby impeding social support, which has been linked to depression, see Ottenbreit & Dobson, 2004).

The findings allowed us to rule out two hypotheses concerning spurious effects. First, the effects of low self-esteem on depression were not confounded by stressful events; that is, low self-esteem and depression are not related simply because challenging life events lead to lower self-esteem and depression, creating a spurious link between the two. Second, reciprocal relations between stressful events and depression were not confounded by low self-esteem; that is, stressful events and depression are not related simply because low self-esteem individuals are prone to depression and tend to experience more stressful events, creating a spurious correlation between the two.

The results of the present research provide a systematic picture of the prospective relations among the three constructs: low self-esteem and stressful events have independent, non-interactive effects on subsequent depression, and depression has an effect on subsequent occurrence of stressful events.

The present research replicates previous studies on the relationship between stressful events and depression (cf. Cole et al., 2006; Hammen, 1991; Holahan et al., 2005), and strengthens conclusions from previous research by showing that the reciprocal relation between stressful events and depression holds after controlling for self-esteem. In addition, the present research suggests that, in contrast to depression, low self-esteem is not a factor of stress generation.

Importantly, the study designs do not allow for strong conclusions regarding the causality of the effects. As in all passive observational designs, effects between factors may be caused by third variables that were not assessed (Finkel, 1995). Nevertheless, longitudinal analyses are

useful because they can indicate whether the data are consistent with a causal model of the relation between the variables.

Also, the results do not allow for firm conclusions with regard to the clinical category of major depressive disorder (MDD). First, the depression measures used in the present research rely on self-report, but conclusions about the antecedents of MDD should be based on diagnoses of depression from clinical interviews. Second, our analyses are based on nonclinical samples, which do not allow for valid conclusions about depressive episodes in clinical populations. Future research, therefore, should test the potential buffering effect of self-esteem on the onset or recurrence of depressive episodes in MDD following stressful events. Nevertheless, the means and standard deviations on the depression measure used, i.e., the CES-D, imply that a fairly substantial proportion of the participants in each of the three studies reported experiencing many of the depressive symptoms at least *some or a little of the time* or *one to two days* per week. Therefore, we believe that the results are relevant for levels of depressive affect that represent a significant impairment in the individual's psychological well-being. Clinically significant levels of depressed mood do not necessarily have to meet the criteria for MDD as given in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev., American Psychiatric Association, 2000) and many persons not in treatment can have clinically significant levels of depressed mood.

Another possible limitation of the present research is that most participants had self-esteem scores that were moderate to high in absolute value; only a small percentage of participants had scores that were below the midpoint of the response scale (specifically, 13%, 11%, and 2% in Studies 1, 2, and 3, respectively). Thus, we do not know whether the present findings generalize to samples of participants who have low self-esteem in an absolute sense.

However, it is important to note that, at least in samples from Western cultures, self-esteem scores tend to be distributed predominantly in the middle to high range (cf. Heine, Lehman, Markus, & Kitayama, 1999; Robins et al., 2002). In other words, individuals who rate their self-esteem at about the midpoint of the response scale actually have low self-esteem relative to the population. Importantly, in all three studies, the self-esteem means and standard deviations were similar to those found in normative population samples. For example, in Study 2, means and standard deviations for the German version of the RSE are similar to normative data from a representative sample, as reported by Roth, Decker, Herzberg, and Brähler (2008); moreover, only about 4% of the representative sample scored below the midpoint of the response scale. Our Study 3 even provides normative data by itself, given that it uses a large probability sample. When the means and standard deviations of Study 3 are adjusted to the different response scale used in Study 1, the distributional characteristics in Study 1 correspond closely to the representative data from Study 3. Given that the distributions of self-esteem scores in our studies are similar to distributions in normative samples, the results of the present research should generalize to normal populations within the age group studied, and therefore it seems warranted to describe the findings in terms of "low" vs. "high" self-esteem.[12]

Another limitation is that we focused on one developmental stage, specifically adolescence and young adulthood. Future research, therefore, should test whether the results hold at other developmental stages, such as midlife and old age. However, as outlined in the introduction, samples with adolescents and young adults are particularly suited to test hypotheses concerning the relations among low self-esteem, stressful events, and depression, suggesting that the present research provides a conservative test of the hypotheses.

A strength of the present research is the convergence of findings across the three studies, despite different study characteristics, increasing confidence in the generalizability of the findings. The studies differed in type of sample (college students in Study 1, employee trainees in Study 2, and a national probability sample of adolescents and young adults in Study 3), nationality (American in Studies 1 and 3, and Swiss in Study 2), length of prospective time intervals (one year in Study 1, six weeks in Study 2, and two years in Study 3), type of stressful events analyzed (stressful life events in Studies 1 and 3, and daily hassles in Study 2), and method to assess stressful events (checklists in Studies 1 and 3, and experience sampling in Study 2). Moreover, Study 2 addressed a limitation of Studies 1 and 3, by assessing stressful events at measurement occasions that were separate from measurement occasions for self-esteem and depression.

Future studies on the self-esteem buffering hypothesis might assess the characteristics of stressful events in more detail. Self-esteem might have a buffering effect only for specific subtypes of stressful events. For example, the literature on stress generation distinguishes between so-called dependent and independent events (i.e., life events that are, or are not, under the individual's control, Hammen, 2005). In the present research, mostly dependent events were assessed: in Study 1, only one of the events was clearly independent (death or serious illness/injury of a close family member or friend), and none of the events in Studies 2 and 3. Therefore, in the present research, no reliable test of differences between dependent and independent events was possible. Future studies could use measures of stressful events that assess a larger set of stressful events, and test the moderating effects of differing event characteristics.[13]

In conclusion, the present research suggests that low self-esteem and stressful events are independent risk factors for depression. These findings have implications for interventions aimed at preventing depression: improving self-esteem reduces risk of depression regardless of whether the individual is enduring stressful or non-stressful life experiences. At the same time, preventing stressful events and improving efficacy to cope with stressful events reduces risk of depression not only among individuals with low self-esteem, but also among high self-esteem individuals.

References

Abela, J. R. Z. (2002). Depressive mood reactions to failure in the achievement domain: A test of the integration of the hopelessness and self-esteem theories of depression. *Cognitive Therapy and Research, 26*, 531-552.

Abela, J. R. Z., & Skitch, S. A. (2007). Dysfunctional attitudes, self-esteem, and hassles: Cognitive vulnerability to depression in children of affectively ill parents. *Behaviour Research and Therapy, 45*, 1127-1140.

Abela, J. R. Z., Webb, C. A., Wagner, C., Ho, M. H. R., & Adams, P. (2006). The role of self-criticism, dependency, and hassles in the course of depressive illness: A multiwave longitudinal study. *Personality and Social Psychology Bulletin, 32*, 328-338.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.)*. Washington, DC: Author.

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Harper and Row.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305-314.

Brown, G. W., & Harris, T. (1978). *Social origins of depression: A study of psychiatric disorder*. New York: Free Press.

Butler, A. C., Hokanson, J. E., & Flynn, H. A. (1994). A comparison of self-esteem lability and low trait self-esteem as vulnerability factors for depression. *Journal of Personality and Social Psychology, 66*, 166-177.

Center for Human Resource Research. (2006). *NLSY79 child and young adult data users guide*. Columbus, OH: Ohio State University.

Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality, 59*, 143-178.

Cheng, S. K., & Lam, D. J. (1997). Relationship among life stress, problem solving, self-esteem, and dysphoria in Hong Kong adolescents: Test of a model. *Journal of Social and Clinical Psychology, 16*, 343-355.

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

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558-577.

Cole, D. A., Nolen-Hoeksema, S., Girgus, J., & Paul, G. (2006). Stress exposure and stress generation in child and adolescent depression: A latent trait-state-error approach to longitudinal analyses. *Journal of Abnormal Psychology, 115*, 40-51.

Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods, 2*, 447-452.

Costello, E. J., Erkanli, A., & Angold, A. (2006). Is there an epidemic of child or adolescent depression? *Journal of Child Psychology and Psychiatry, 47*, 1263-1271.

Davila, J., Bradbury, T. N., Cohan, C. L., & Tochluk, S. (1997). Marital functioning and depressive symptoms: Evidence for a stress generation model. *Journal of Personality and Social Psychology, 73*, 849-861.

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological Science, 16*, 328-335.

Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Volume 3 (Instruments for adults)* (pp. 363-377). Mahwah, NJ: Erlbaum.

Fernandez, M. E., Mutran, E., & Reitzes, D. C. (1998). Moderating the effects of stress on depressive symptoms. *Research on Aging, 20*, 163-182.

Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage.

Fraley, R. C., & Marks, M. J. (2007). The null hypothesis significance-testing debate and its implications for personality research. In R. W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 149-169). New York: Guilford.

Galambos, N. L., Barker, E. T., & Krahn, H. J. (2006). Depression, self-esteem, and anger in emerging adulthood: Seven-year trajectories. *Developmental Psychology, 42*, 350-365.

Grant, K. E., Compas, B. E., Thurm, A. E., McMahon, S. D., & Gipson, P. Y. (2004). Stressors and child and adolescent psychopathology: Measurement issues and prospective effects. *Journal of Clinical Child and Adolescent Psychology, 33*, 412-425.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1-20.

Hammen, C. (1991). Generation of stress in the course of unipolar depression. *Journal of Abnormal Psychology, 100*, 555-561.

Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology, 1*, 293-319.

Hankin, B. L., Fraley, R. C., Lahey, B. B., & Waldman, I. D. (2005). Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *Journal of Abnormal Psychology, 114*, 96-110.

Hautzinger, M., & Bailer, M. (1993). *Allgemeine Depressions-Skala (ADS): Manual [Center for Epidemiologic Studies Depression Scale (CES-D): Manual]*. Weinheim: Beltz.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*, 766-794.

Holahan, C. J., Moos, R. H., Holahan, C. K., Brennan, P. L., & Schutte, K. K. (2005). Stress generation, avoidance coping, and depressive symptoms: A 10-year model. *Journal of Consulting and Clinical Psychology, 73*, 658-666.

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.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin, 117*, 348-357.

Kernis, M. H., Whisenhunt, C. R., Waschull, S. B., Greenier, K. D., Berry, A. J., Herlocker, C. E., et al. (1998). Multiple facets of self-esteem and their relations to depressive symptoms. *Personality and Social Psychology Bulletin, 24*, 657-668.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*, 593-602.

Kim, K. J., Conger, R. D., Elder, G. H., & Lorenz, F. O. (2003). Reciprocal influences between stressful life events and adolescent internalizing and externalizing problems. *Child Development, 74*, 127-143.

Lazarus, R. S. (1999). *Stress and emotion: A new synthesis*. New York: Springer.

Lewinsohn, P. M., Hoberman, H. M., & Rosenbaum, M. (1988). A prospective study of risk factors for unipolar depression. *Journal of Abnormal Psychology, 97*, 251-264.

Lewinsohn, P. M., Solomon, A., Seeley, J. R., & Zeiss, A. (2000). Clinical implications of "subthreshold" depressive symptoms. *Journal of Abnormal Psychology, 109*, 345-351.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151-173.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201-226.

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods, 11*, 19-35.

Metalsky, G. I., Joiner, T. E., Hardin, T. S., & Abramson, L. Y. (1993). Depressive reactions to failure in a naturalistic setting: A test of the hopelessness and self-esteem theories of depression. *Journal of Abnormal Psychology, 102*, 101-109.

Mirowsky, J., & Kim, J. (2007). Graphing age trajectories: Vector graphs, synthetic and virtual cohort projections, and cross-sectional profiles of depression. *Sociological Methods and Research, 35*, 497-541.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (2000). Self-esteem and the quest for felt security: How perceived regard regulates attachment processes. *Journal of Personality and Social Psychology, 78*, 478-498.

Murray, S. L., Rose, P., Bellavia, G. M., Holmes, J. G., & Kusche, A. G. (2002). When rejection stings: How self-esteem constrains relationship-enhancing processes. *Journal of Personality and Social Psychology, 83*, 556-573.

Murrell, S. A., Meeks, S., & Walker, J. (1991). Protective functions of health and self-esteem against depression in older adults facing illness or bereavement. *Psychology and Aging, 6*, 352-360.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide: Fifth edition*. Los Angeles, CA: Muthén and Muthén.

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology, 109*, 504-511.

Oatley, K., & Bolton, W. (1985). A social-cognitive theory of depression in reaction to life events. *Psychological Review, 92*, 372-388.

Ormel, J., Oldehinkel, A. J., & Vollebergh, W. (2004). Vulnerability before, during, and after a major depressive episode. *Archives of General Psychiatry, 61*, 990-996.

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology, 95*, 695-708.

Orth, U., Robins, R. W., Trzesniewski, K. H., Maes, J., & Schmitt, M. (2008). *Low self-esteem is a risk factor for depressive symptoms from young adulthood to old age*. Manuscript under review.

Ottenbreit, N. D., & Dobson, K. S. (2004). Avoidance and depression: The construction of the Cognitive-Behavioral Avoidance Scale. *Behaviour Research and Therapy, 42*, 293-313.

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research, 39*, 303-328.

Prisciandaro, J. J., & Roberts, J. E. (2005). A taxometric investigation of unipolar depression in the national comorbidity survey. *Journal of Abnormal Psychology, 114*, 718-728.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Ralph, J. A., & Mineka, S. (1998). Attributional style and self-esteem: The prediction of emotional distress following a midterm exam. *Journal of Abnormal Psychology, 107*, 203-215.

Roberts, J. E. (2006). Self-esteem from a clinical perspective. In M. H. Kernis (Ed.), *Self-esteem issues and answers: A sourcebook of current perspectives* (pp. 298-305). New York: Psychology Press.

Roberts, J. E., & Gotlib, I. H. (1997). Temporal variability in global self-esteem and specific self-evaluation as prospective predictors of emotional distress: Specificity in predictors and outcome. *Journal of Abnormal Psychology, 106*, 521-529.

Roberts, J. E., & Monroe, S. M. (1992). Vulnerable self-esteem and depressive symptoms: Prospective findings comparing three alternative conceptualizations. *Journal of Personality and Social Psychology, 62*, 804-812.

Roberts, J. E., & Monroe, S. M. (1999). Vulnerable self-esteem and social processes in depression: Toward an interpersonal model of self-esteem regulation. In T. Joiner & J. C.

Coyne (Eds.), *The interactional nature of depression: Advances in interpersonal approaches* (pp. 149-187). Washington, DC: American Psychological Association.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*, 151-161.

Robins, R. W., Noftle, E. E., Trzesniewski, K. H., & Roberts, B. W. (2005). Do people know how their personality has changed? Correlates of perceived and actual personality change in young adulthood. *Journal of Personality, 73*, 481-521.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17*, 423-434.

Robinson, N. S., Garber, J., & Hilsman, R. (1995). Cognitions and stress: Direct and moderating effects on depressive versus externalizing symptoms during the junior high school transition. *Journal of Abnormal Psychology, 104*, 453-463.

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

Roth, M., Decker, O., Herzberg, P. Y., & Brähler, E. (2008). Dimensionality and norms of the Rosenberg Self-Esteem Scale in a German general population sample. *European Journal of Psychological Assessment, 24*, 190-197.

Ruscio, J., & Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology, 109*, 473-487.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Southall, D., & Roberts, J. E. (2002). Attributional style and self-esteem in vulnerability to adolescent depressive symptoms following life stress: A 14-week prospective study. *Cognitive Therapy and Research, 26*, 563-579.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment, 80*, 217-222.

Tram, J. M., & Cole, D. A. (2000). Self-perceived competence and the relation between life events and depressive symptoms in adolescence: Mediator or moderator? *Journal of Abnormal Psychology, 109*, 753-760.

von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl nach Rosenberg [A revised German version of the Rosenberg Self-Esteem Scale]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 24*, 3-7.

Whisman, M. A., & Kwon, P. (1993). Life stress and dysphoria: The role of self-esteem and hopelessness. *Journal of Personality and Social Psychology, 65*, 1054-1060.

Author Note

Footnotes

[1] We use the term "depression" to denote a continuous variable (i.e., individual differences in depressive affect), rather than a clinical category such as major depressive disorder. Taxometric analyses suggest that depression is best conceptualized as a continuous construct (Hankin, Fraley, Lahey, & Waldman, 2005; Lewinsohn, Solomon, Seeley, & Zeiss, 2000; Prisciandaro & Roberts, 2005; Ruscio & Ruscio, 2000).

[2] These data were used in a previous study on the relation between self-esteem and depression (Orth, Robins, & Roberts, 2008); however, that study did not include any analyses of stressful events.

[3] Given the high number of dimensions of integration in Studies 1 and 3 (i.e., six dimensions of integration), we could not employ the default algorithm, i.e., rectangular integration, but used the Monte Carlo algorithm, following the recommendation by Muthén and Muthén (2007). For reasons of consistency across studies, we used the Monte Carlo algorithm also in Study 2.

[4] Grant et al. (2004) report test-retest reliability estimates for several stressful event checklists ranging from .74 to .96. We therefore decided to use .80 as a reliability estimate for the stressful event measures used in the present research.

[5] In Study 1, as well as in Studies 2 and 3, the measure of stressful events was not normally distributed. We therefore tested whether using logarithmically transformed measures of stressful events yielded different results compared to using the original scaling: in all three studies, the results were very similar and did not lead to different conclusions, in particular with regard to the interaction effects, and all of the significant paths remained significant, and the nonsignificant interaction effects remained nonsignificant.

[6] In addition to testing for gender differences simultaneously for all structural coefficients

we inspected gender differences in individual structural coefficients in all three studies. There

were no gender differences that replicated across studies, except that the effect of stressful events

on depression was stronger for women than for men (in Study 1, the average coefficients for

female and male participants were .23 and .13, respectively; in Study 2, .20 and .13, respectively;

and in Study 3, .14 and .09, respectively). We also tested for ethnicity differences in the

structural coefficients (Asians vs. others) in Study 1, using a multiple group analysis. A model

allowing for different coefficients for Asians vs. other ethnicities did not significantly improve

model fit, relative to a model with constraints across ethnicities; the difference in model fit was

$\Delta\chi^2(13, N = 359) = 18.1$, *ns*.

[7] We also tested two further mediation hypotheses. First, depression had an indirect effect

on subsequent depression mediated by intermediate stressful events, as indicated by a significant

Sobel Test. This indirect effect was also significant in Study 3, but failed to replicate in Study 2.

Thus, the temporal stability of depression is partially due to the reciprocal relationship between

depression and stressful events. Second, self-esteem had an indirect effect on stressful events

mediated by depression, as indicated by a significant Sobel Test. This indirect effect was also

significant in Study 3 (it could not be tested in Study 2 because only two waves of data were

available). Thus, even if the direct of self-esteem on stressful events is very small, self-esteem

indirectly influences the occurrence of subsequent stressful events by its effect on depression.

[8] Study 2 also allowed to test whether the stability (rather than the level) of self-esteem

buffers the effects of stressful events on depression, as some researchers have argued (cf. Kernis

et al., 1998; Roberts & Gotlib, 1997). At the same measurement occasions as stressful events

were assessed, participants completed a 5-item version of the RSE, which was adapted to

measure momentary self-esteem (alpha = .94). We used the intraindividual standard deviation of self-esteem across time as an indicator of (low) self-esteem stability. However, the results showed that self-esteem stability did not interact with stressful events in the prediction of subsequent depression.

[9] These data were used in a previous study on the relation between self-esteem and depression (Orth, Robins, & Roberts, 2008); however, that study did not include any analyses of stressful events.

[10] The third and fourth item was assessed for the past 12 months.

[11] For Time 4, the first two items were not included in the computation of alpha reliability, because the items were not applicable for most of the participants between age 19 and 21, resulting in a large number of missings for these items at Time 4.

[12] We examined the scatterplot of residuals (plotted against self-esteem scores) from the regression of depression on self-esteem and they showed an essentially random distribution in all three studies, suggesting that the relation between self-esteem and depression is similar across all levels of self-esteem. To further illustrate this point, we divided the samples into self-esteem quintiles and examined mean depression scores in each quintile. These means show that depression levels declined in a fairly linear manner from the bottom self-esteem quintile (representing the 20% of participants with the lowest self-esteem levels) to the top self-esteem quintile (representing the top 20% of participants). For example, in Study 1, the mean depression scores were 1.32, 1.08, 0.87, 0.75, and 0.48, for self-esteem quintiles ordered from low to high.

[13] One could argue that self-esteem only serves a buffering effect for events that are self-esteem relevant, such as academic failure. Study 1 included the most extensive measure of stressful events and therefore allowed us to test the self-esteem buffering effect specifically for

academic and competence-related stressful events (5 of the 12 total events). However, no

significant interaction emerged between self-esteem and this subset of events.

Table 1

Summary of Previous Studies Testing the Interactive Effects of Self-Esteem and Stressful Events on Subsequent Depression,

Controlling for Prior Depression

Study	N	% female	Age group	Type of events	Prospective time interval	Self-esteem buffering interaction effect[a]
Abela (2002)	136	53%	adolescents	discrete academic event	8 weeks	+
Abela et al. (2006)	102	86%	adults	daily hassles	6 weeks[b]	3-way
Abela & Skitch (2007)	140	51%	children	daily hassles	6 weeks[b]	3-way
Butler et al. (1994)	73	77%	young adults	stressful life events	5 months	*ns*
Cheng & Lam (1997)	286	27%	adolescents	daily hassles, stressful life events	3 months	*ns*[c]
Fernandez et al. (1998)	729	52%	adults	stressful life events	2 years	+[d]
Kernis et al. (1998)	98	86%	young adults	daily hassles	4-5 weeks	*ns*
Lewinsohn et al. (1988)	354	69%	adults	stressful life events	8 months	*ns*
Metalsky et al. (1993)	114	N/A	young adults	discrete academic event	3-7 days	+
Murrell et al. (1991)	1,074	66%	adults	stressful life events	2 years	*ns*

Ralph & Mineka (1998)	141	54%	young adults	discrete academic event	10-13 days	+
Roberts & Monroe (1992)	192	64%	young adults	discrete academic event	7-9 days	*ns*
Robinson et al. (1995)	381	58%	adolescents	daily hassles, stressful life events[e]	4-5 months	3-way
Southall & Roberts (2002)	115	50%	adolescents	stressful life events	3 months	*ns*
Whisman & Kwon (1993)	80	66%	young adults	daily hassles, stressful life events[e]	3 months	-

[a] "+" denotes a significant two-way interaction effect consistent with the self-esteem buffering hypothesis; "-" denotes a significant two-way interaction effect contrary to the self-esteem buffering hypothesis; "3-way" denotes a significant three-way interaction effect of self-esteem, stressful events, and a third variable, while the two-way interaction effect of self-esteem and stressful events was nonsignificant; and "*ns*" denotes a nonsignificant two-way interaction effect in absence of any significant three-way interaction effect involving self-esteem and stressful events.

[b] The data included multiple waves over one year, separated by six-week intervals. Effects were estimated by multilevel modeling.

[c] The regression equation included two-way interactions for both daily hassles and stressful life events.

[d] Stressful life events were divided into two scales: social network events and work disruptions. The interaction of self-esteem and stressful life events was significant for social network events, but not work disruptions.

[e] The analyses were conducted for a variable combining daily hassles and stressful life events.

Table 2

Means and Standard Deviations of Measures (Study 1)

Variable	Time 1 (18 years)		Time 2 (19 years)		Time 3 (20 years)		Time 4 (21 years)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
RSE	3.82	0.77	3.88	0.82	3.86	0.77	4.06	0.72
Stressful events[a]	--	--	1.41	1.30	1.22	1.17	1.26	1.36
CES-D	0.98	0.58	0.94	0.57	0.82	0.53	0.74	0.52

Note. Response scales ranged from 1 to 5 for the RSE and from 0 to 3 for the CES-D. The range of possible values for stressful events was 0 to 12. RSE = Rosenberg Self-Esteem Scale; CES-D = Center for Epidemiologic Studies Depression Scale.

[a] Stressful events were assessed for the time interval that preceded the assessment. For example, the score at Time 2 indicates the number of stressful events between Time 1 and 2.

Table 3

Fit Indices of the Models Tested (Studies 1 to 3)

Model	χ^2	*df*	TLI	CFI	RMSEA (90%-CI)	BIC
Study 1 (*N* = 359)						
No Interaction	377.2*	290	.98	.98	.029 (.020-.037)	10497.7
Unconstrained Interaction	--	--	--	--	--	10519.3
Constrained Interaction	--	--	--	--	--	10515.2
Study 2 (*N* = 249)						
No Interaction	116.7*	54	.96	.97	.070 (.053-.088)	3964.5
Interaction	--	--	--	--	--	3969.9
Study 3 (*N* = 2,403)						
No Interaction	891.2*	290	.93	.94	.030 (.027-.032)	60573.8
Unconstrained Interaction	--	--	--	--	--	60624.0
Constrained Interaction	--	--	--	--	--	60609.2

Note. For models including interactions, χ^2, *df*, TLI, CFI, and RMSEA are not available. TLI =

Tucker-Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of

Approximation; CI = confidence interval; BIC = Bayesian Information Criterion.

* $p < .05$.

Table 4

Unstandardized Estimates of Structural Coefficients (Study 1)

Coefficients	No Interaction Model	Unconstrained Interaction Model	Constrained Interaction Model
Effects on self-esteem			
Self-esteem	0.83 (0.03)*	0.83 (0.04)*	0.83 (0.04)*
Stressful events	-0.02 (0.01)	-0.02 (0.02)	-0.02 (0.02)
Depression	0.01 (0.04)	0.01 (0.05)	0.01 (0.04)
Effects on stressful events			
Self-esteem	0.14 (0.11)	0.16 (0.13)	0.14 (0.12)
Stressful events	0.56 (0.06)*	0.58 (0.07)*	0.58 (0.07)*
Depression	0.41 (0.13)*	0.42 (0.15)*	0.41 (0.15)*
Effects on depression			
Self-esteem	-0.22 (0.05)*	-0.24 (0.06)*	-0.23 (0.06)*
Stressful events	0.08 (0.02)*	0.08 (0.02)*	0.08 (0.02)*
Depression	0.28 (0.06)*	0.29 (0.08)*	0.28 (0.08)*
IA Time 1 to 2	--	-0.06 (0.06)	0.00 (0.05)
IA Time 2 to 3	--	0.17 (0.08)*	0.00 (0.05)
IA Time 3 to 4	--	-0.08 (0.10)	0.00 (0.05)

Note. Standard errors of estimates are given in parentheses. IA = latent interaction between self-esteem and stressful events.

* $p < .05$.

Table 5

Means and Standard Deviations of Measures (Study 2)

	Time 1		Time 2		Time 3	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
RSE	3.78	0.86	--	--	3.74	0.90
Stressful events per day	--	--	0.30	0.43	--	--
CES-D	0.68	0.56	--	--	0.72	0.59

Note. Response scales ranged from 0 to 5 for the RSE and from 0 to 3 for the CES-D. The range of possible values for stressful events per day was 0 to 5. RSE = Rosenberg Self-Esteem Scale; CES-D = Center for Epidemiologic Studies Depression Scale.

Table 6

Unstandardized Estimates of Structural Coefficients (Study 2)

Coefficients	No Interaction Model	Interaction Model
Effects on self-esteem		
Self-esteem	0.92 (0.10)*	0.93 (0.10)*
Stressful events	-0.21 (0.13)	-0.22 (0.13)
Depression	0.01 (0.12)	0.02 (0.12)
Effects on stressful events		
Self-esteem	0.11 (0.10)	0.11 (0.09)
Depression	0.23 (0.11)*	0.24 (0.11)*
Effects on depression		
Self-esteem	-0.36 (0.11)*	-0.37 (0.11)*
Stressful events	0.28 (0.13)*	0.29 (0.13)*
Depression	0.35 (0.16)*	0.34 (0.16)*
IA	--	-0.04 (0.15)

Note. Standard errors of estimates are given in parentheses. IA = latent interaction between self-esteem and stressful events.

* $p < .05$.

Table 7

Means and Standard Deviations of Measures (Study 3)

	Time 1 (15 years)		Time 2 (17 years)		Time 3 (19 years)		Time 4 (21 years)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
RSE	3.19	0.41	3.25	0.42	3.29	0.45	3.27	0.42
Stressful events[a]	--	--	0.72	0.96	0.63	0.93	0.60	1.05
CES-D	0.71	0.52	0.69	0.51	0.70	0.54	0.68	0.54

Note. Response scales ranged from 1 to 4 for the RSE and from 0 to 3 for the CES-D. The range of possible values for stressful events was 0 to 5. RSE = Rosenberg Self-Esteem Scale; CES-D = Center for Epidemiologic Studies Depression Scale.

[a] Stressful events were assessed for the time interval that preceded the assessment. For example, the score at Time 2 indicates the number of stressful events between Time 1 and 2.

Table 8

Unstandardized Estimates of Structural Coefficients (Study 3)

Coefficients	No Interaction Model	Unconstrained Interaction Model	Constrained Interaction Model
Effects on self-esteem			
Self-esteem	0.61 (0.02)*	0.62 (0.03)*	0.62 (0.03)*
Stressful events	-0.01 (0.01)*	-0.01 (0.01)*	-0.01 (0.01)*
Depression	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)
Effects on stressful events			
Self-esteem	-0.09 (0.05)*	-0.09 (0.05)*	-0.09 (0.05)*
Stressful events	0.31 (0.02)*	0.31 (0.03)*	0.31 (0.03)*
Depression	0.22 (0.05)*	0.22 (0.05)*	0.22 (0.05)*
Effects on depression			
Self-esteem	-0.11 (0.03)*	-0.12 (0.04)*	-0.12 (0.04)*
Stressful events	0.05 (0.01)*	0.05 (0.01)*	0.05 (0.01)*
Depression	0.50 (0.03)*	0.49 (0.04)*	0.49 (0.04)*
IA Time 1 to 2	--	-0.05 (0.08)	-0.07 (0.04)
IA Time 2 to 3	--	-0.14 (0.09)	-0.07 (0.04)
IA Time 3 to 4	--	-0.05 (0.08)	-0.07 (0.04)

Note. Standard errors of estimates are given in parentheses. IA = latent interaction between self-esteem and stressful events.

* $p < .05$.

Figure Captions

Figure 1. Alternative hypotheses of the relations between self-esteem, stressful events, and depression: self-esteem buffering hypothesis, i.e., self-esteem buffers the effects of stressful events on depression (Hypothesis 1); stressful events mediate the effects of low self-esteem on depression (Hypothesis 2); low self-esteem mediates the effects of stressful events on depression (Hypothesis 3); stressful events account for the relation between self-esteem and depression (Hypothesis 4); and self-esteem accounts for the relation between stressful events and depression (Hypothesis 5).

Figure 2. The figure illustrates a structural model of the relations among self-esteem, stressful events, and depression. Latent interactions are symbolized by a filled circle.

Figure 3. Standardized structural coefficients for the "no interaction" model (Study 1). The figure shows only latent constructs and omits observed variables. * $p < .05$.

Figure 4. Standardized structural coefficients for the "no interaction" model (Study 2). The figure shows only latent constructs and omits observed variables. * $p < .05$.

Figure 5. Standardized structural coefficients for the "no interaction" model (Study 3). The figure shows only latent constructs and omits observed variables. * $p < .05$.
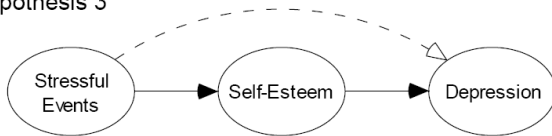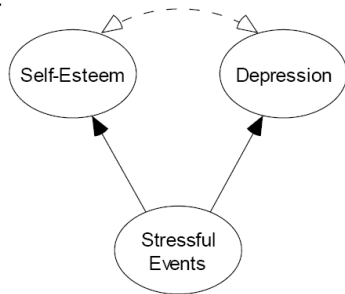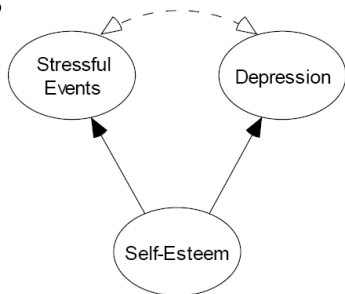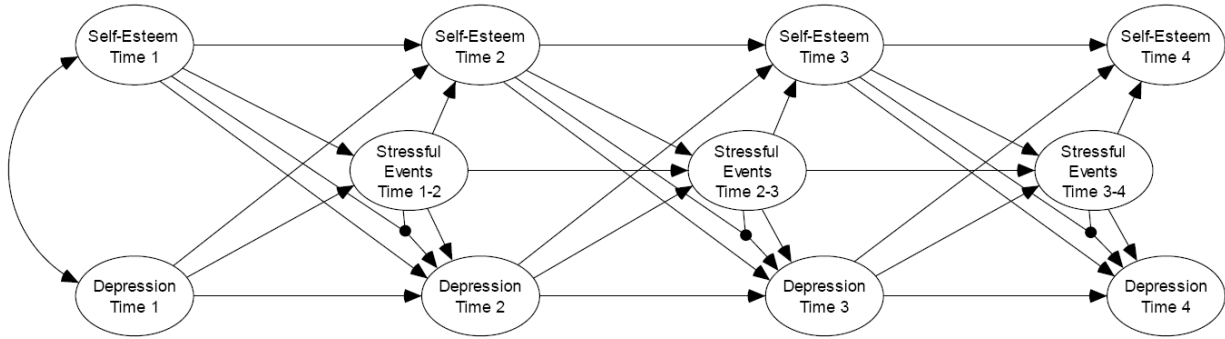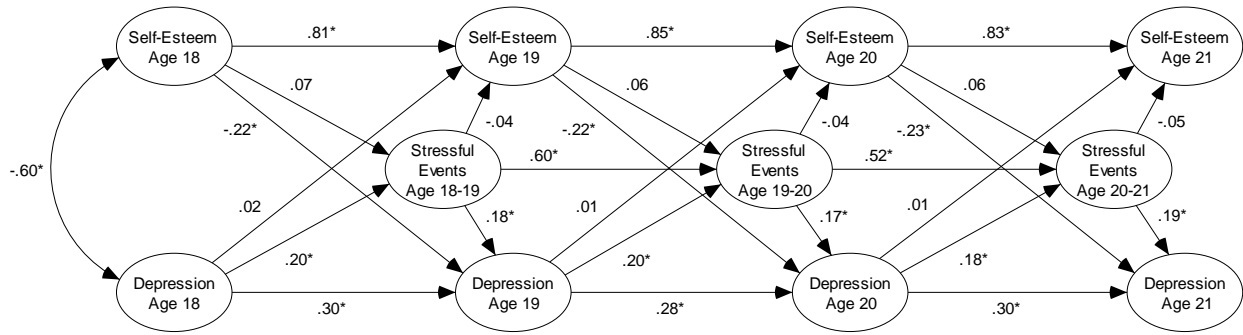
Hypothesis 1
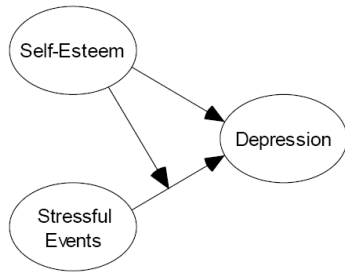
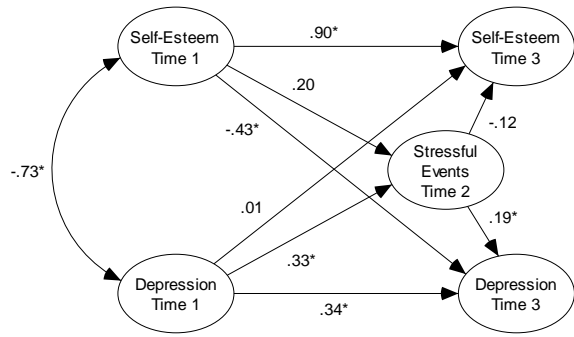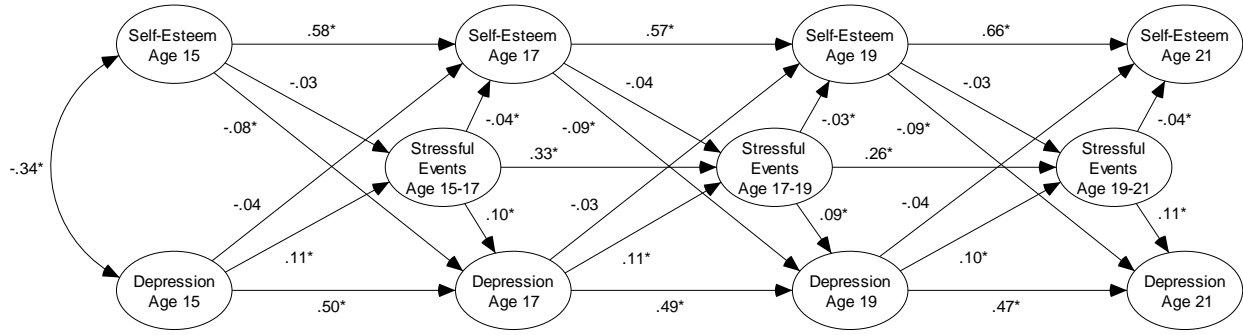Hypothesis 2
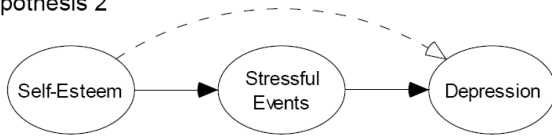
Hypothesis 3

Hypothesis 4

Hypothesis 5

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5