Psychology and Aging
2011, Vol. 26, No. 4, 987–993

0882-7974/11/$12.00 DOI: 10.1037/a0024406

BRIEF REPORT

Aging and the Structure and Long-Term Stability of the Internalizing Spectrum of Personality and Psychopathology

Nicholas R. Eaton and Robert F. Krueger
University of Minnesota

Thomas F. Oltmanns
Washington University in St. Louis

Structural psychopathology research has identified two broad factors—internalizing and externalizing—that account for comorbidity among many common mental disorders. Evaluating the utility of these factors for nosology, research, and treatment entails expanding beyond a cross-sectional understanding to how these factors evolve over time. We tested factorial invariance of internalizing in three age cohort groups—35 years and under (n = 1,729), 36–50 years (n = 2,719), and over 50 years (n = 2,601)—as well as the long-term stability of internalizing within individuals. Internalizing showed a notable degree of invariance between cohorts and within cohorts over time; long-term internalizing stability was equivalently moderate-to-high in each cohort.

Keywords: aging, internalizing, psychopathology, neuroticism, stability, invariance

Comorbid presentations of putatively distinct forms of mental disorder, such as major depression (MD) and generalized anxiety disorder (GAD), occur with greater frequency than expected by chance alone, which suggests that multiple diagnoses may reflect the same underlying core liabilities (Watson, 2009). Multivariate statistical analyses have demonstrated that the comorbidity of many common mood and anxiety disorders can be well captured by a broad latent liability factor called internalizing (INT; Krueger & Markon, 2006). INT accounts for comorbidity of disorders such as MD, GAD, panic disorder (PAN), dysthymia, social phobia, and specific phobia as well as their links to trait neuroticism (NEUR; Eaton, South, & Krueger, 2010; Hettema, Neale, Myers, Prescott, & Kendler, 2006; Griffith et al., 2010).

In addition to accounting for comorbidity, INT has emerged as a crucial construct in understanding public health generally and psychopathology specifically. Higher levels of INT and its indicators, such as neuroticism, are associated with numerous poor health outcomes, such as physical disorders, utilization of health care resources, and perhaps the overall "quality and longevity of our lives" (Lahey, 2009, p. 241). Etiologically, INT reflects genotypic covariance of many mood and anxiety disorders in addition to phenotypic covariance (Kendler et al., 2011) and INT is the primary pathway by which mood and anxiety disorders originate and recur over time (Kessler et al., 2011). Given its utility for understanding much emotion-based psychopathology, INT also has implications for nosology: DSM-5 and ICD-11 mood and anxiety disorders will likely be reorganized into an INT "meta-structure" (rather than current "mood disorders" and "anxiety disorders" chapters; see Andrews et al., 2009; Regier, Narrow, Kuhl, & Kupfer, 2011). Finally, interventions targeting the INT liability of emotion-based psychopathology, rather than its specific manifestations, are a major focus of contemporary psychotherapy research (Barlow et al., 2011).

Structure and Stability of INT in the Context of Aging

While the utility of INT in multiple domains has been illustrated, the construct has not been formally investigated with regard to aging. Thus, two critical questions remain unanswered. The first question is one of factorial invariance: Is the structure of INT invariant as individuals age (i.e., within-person developmental change) and across age cohorts (i.e., between-person age-cohort differences)? While INT has been shown to be largely invariant across gender (Kramer, Krueger, & Hicks, 2008) and individuals from various nations (Krueger, Chentsova-Dutton, Markon, Goldbert, & Ormel, 2003), no studies have examined INT invariance in an aging context. This uncertainty regarding INT age invariance has been identified as a fundamental gap in our understanding of internalizing (Eaton et al., 2010), especially given what some authors have referred to as a looming "geriatric crisis" in mental health care due to rapidly aging populations in the United States and abroad (Jeste et al., 1999; Horn & McArdle, 1992).

The potential age invariance of INT is promising given the reported invariance of several related, nondiagnostic constructs,

This article was published Online First July 4, 2011.

Nicholas R. Eaton and Robert F. Krueger, Department of Psychology, University of Minnesota; and Thomas F. Oltmanns, Department of Psychology, Washington University in St. Louis.

This research was supported in part by a fellowship awarded to the first author by the Summer Training on Aging Research Topics—Mental Health (START-MH) program, conducted by the Division of Geriatric Psychiatry at the University of California, San Diego, with grant support from the National Institute of Mental Health. The MIDUS study was supported by the John D. and Catherine T. MacArthur Foundation Research Network on Successful Midlife Development and by National Institute on Aging Grant AG20166.

Correspondence concerning this article should be addressed to Nicholas R. Eaton, N218 Elliott Hall, 75 East River Road, Minneapolis, MN 55455-0344. E-mail: nreaton@gmail.com

such as trait neuroticism (Allemand, Zimprich, & Hertzog, 2007) and anxious-arousal (Teachman, Siedlecki, & Magee, 2007), but explicit investigation of INT is necessary to clarify its within-person developmental, and between-person age-cohort, structure and to realize fully its potential for public health, nosology, and treatment. Further, invariance investigations of INT have the potential to inform developmental theories of psychopathology and personality. For instance, prevalence rates of mood and anxiety disorders tend to be lower in older individuals (Kessler et al., 1994; Regier et al., 1988); without the establishment of age invariance, such comparisons may reflect putatively similar manifestations (disorders) of different constructs (non-age invariant INT) and thus should be interpreted with some caution. In addition, such analyses can test hypotheses regarding emotional experiences and aging. Some theories posit that emotional experience is more similar across adult development (e.g., Lawton, Kleban, Rajagopal, & Dean, 1992), in which case one might expect INT invariance across cohorts and over time within individuals; others suggest that there are age-related increases in emotional maturity and health (e.g., Carstensen, Gross, & Fung, 1997; Isaacowitz, Charles, & Carstensen, 2000), in which case one might expect mean/intercept declines across cohorts and over time within individuals, or even lack of invariance. Inferences drawn from such analyses will depend on the definition of cohorts to some extent, as studies comparing broadly defined cohorts address these issues at a different level than studies comparing temporally more constricted cohorts.

The second question that requires attention regards stability: How stable is the INT liability within individuals over time? This question is particularly important given INT's key role in the maintenance and emergence of many mental disorders over time (Kessler et al., 2011). The few studies that have addressed this issue in adults (Fergusson, Horwood, & Boden, 2006; Krueger, Caspi, Moffitt, & Silva, 1998; Vollebergh et al., 2001) indicated moderate-to-high levels of short-term INT stability within individuals. These studies focused on short-term (e.g., 1-year) stability and/or only young adults (typically ages 18–21 or 21–25), and they did not evaluate whether INT stability was similar in different age cohorts; thus, formal investigation of long-term stability of INT is necessary, as is testing whether this stability is moderated by age cohort.

The Present Study

The present study examined the factorial invariance and long-term stability of INT in an aging context. We investigated two separate questions. First, to what degree is INT similar in younger, middle, and older cohort age groups? Second, what degree of long-term stability does INT show within individuals over time, and does this stability differ across age cohorts? To address these questions, we analyzed data from a large United States national probability sample collected in two waves over 9 to 10 years.

Method

Participants

We analyzed data from 7,108 individuals (51.1% female), sampled via random digit dialing, who participated in the Midlife in the United States (MIDUS) study of adult development, comprising two data collection waves (MIDUS-1 and MIDUS-2) separated by 9 to 10 years. The MIDUS-2 sample represents the subset of MIDUS-1 participants (n = 4,963) who were successfully recontacted and agreed to participate again. Adjusted for mortality, the overall retention rate was 70%. For more information on sampling and retention, see Radler and Ryff (2010). At MIDUS-1, beginning in 1995, participants' ages ranged from 20 to 75 years (M = 46.4, SD = 13.0; age data missing for 59 individuals), and at the time of MIDUS-2, beginning in 2004, ages ranged from 28 to 84 years (M = 55.4, SD = 12.4). Participants reported their ethnicities as: 78.8% White, 4.5% African American, .8% Asian American, .6% Multiracial, .5% Native American, and 1.7% "other" (13.1% missing). Individuals were compensated $20 for completion of MIDUS-1 and up to $60 for completion of MIDUS-2. The suffix "-1" indicates variables from MIDUS-1 assessment (e.g., MD-1) and "-2" indicates MIDUS-2 (e.g., MD-2).

Assessment

Psychopathological symptoms. *DSM–III–R* MD, GAD, and PAN were assessed for the previous 12-months via the Word Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF; Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998). The CIDI-SF is a structured interview showing high (93%-99%) agreement with the full CIDI. Further, clinical reappraisals demonstrated good agreement with clinician diagnoses (Wang, Berglund, & Kessler, 2000).

Neuroticism. Participants were asked to "Please indicate how well each of the following describes you" by rating four adjectives (moody, worrying, nervous, and calm [reversed]) on a 4-point Likert-type scale. NEUR was calculated as the mean of these ratings, with higher scores indicating higher NEUR. NEUR was considered missing for individuals rating fewer than two adjectives. NEUR scores were available for 6,265 MIDUS-1 participants and 4,000 MIDUS-2 participants. NEUR items formed a reliable scale (α = .74).

Analyses

Exploratory factor analysis was conducted in *R* using an oblique rotation. Scree plot and parallel analysis were used to determine the number of factors to extract (see Brown, 2006). Structural equation modeling was conducted on continuous symptom count data for MD, GAD, PAN, and NEUR in Mplus 5.1 (Muthén & Muthén, 2007). Because symptom counts showed positive skewness, MD, GAD, and PAN were natural log transformed and a maximum likelihood estimator (MLR) robust to non-normality was used. NEUR was approximately normally distributed and not transformed. All analyses accounted for the familial clustering. Full information maximum likelihood was used to account for missingness. In the structural equation models, the variance of the latent INT factor was constrained to one.

Model fit was evaluated with the root mean squared error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and the Bayesian information criterion (BIC). Hu and Bentler (1999) suggested the following cutoffs as indicating reasonably good fit: RMSEA less than .06 and CFI/TLI values of .95 or greater. BIC allows for direct comparison of non-nested

models and balances overall model fit with model parsimony, with lower BIC values indicating the best fitting, most parsimonious model. We defined the optimal model as the model with the lowest BIC that also produced acceptable fit according to the other indices. Satorra-Bentler scaled chi-square difference tests were used to compare parameters in nested models.

Participants were assigned to one of three groups based on their MIDUS-1 ages: a younger cohort (35 years and under, $n = 1{,}729$), a middle cohort (36–50 years, $n = 2{,}719$), and an older cohort (over 50 years, $n = 2{,}601$). These cutoffs best balanced the need for acceptable indicator covariances within groups and produced similar group sizes. At MIDUS-1, these three groups had mean ages (and standard deviations) of 30.4 (3.2), 42.9 (4.3), and 60.7 (6.7), respectively; at MIDUS-2, these groups' ages were 39.7 (3.2), 51.9 (4.3), and 69.0 (6.5), respectively. Cohort labels were selected as a convenient way to describe the relative age standings of the three cohorts. They are *not* intended to indicate that all individuals in, for example, the older cohort are older adults. Cohorts had similar gender compositions: The younger cohort was 52.2% female, the middle was 50.0%, and the older was 52.6%.

Results

Factorial Invariance Analyses

All indicators were significantly ($p < .01$) positively intercorrelated, within and across time points, and showed significant stability over time ($r = .21$ to $.67$), supporting the notion that one or more common factors could account for their shared variance. We next tested factorial invariance using four increasingly stringent levels outlined by Meredith (1993).

Configural invariance. The first, least stringent level of factorial invariance is configural invariance. Configural invariance would be established if the indicators loaded primarily onto only one factor (i.e., INT). We tested this via six exploratory factor analyses—one for each time point across groups (i.e., three age groups × two time points per age group = six analyses). In all cases, scree plot and parallel analyses suggested a one-factor solution, establishing configural invariance and the presence of a single factor across time points in all age groups.

To test more stringent invariance levels, MIDUS-1 and MIDUS-2 data from the three groups were modeled simultaneously. The measurement model consisted of MD-1, GAD-1, PAN-1, and NEUR-1 as indicators for INT-1; MD-2, GAD-2, PAN-2, and NEUR-2 were indicators for INT-2. INT-1 and INT-2 were allowed to covary. Correlated residuals were included between each indicator variable over time (e.g., MD-1 and MD-2), resulting in the structural model depicted in Figure 1. We began by estimating this model with all of its parameters freely estimated, although factor means were set at zero (Model 1). Parameterization descriptions and fit indices for this model, and all subsequent models, are presented in Table 1. Fit indices indicated Model 1 fit the data adequately; however, this model is not parsimonious because parameters are estimated freely across groups and over time points.

Metric invariance. Metric invariance has the next most stringent constraints, wherein the factor loading that links each indicator to INT is constrained to be equal across the age groups and both time points. Models 2, 3, 4, and 5 were precursors to metric

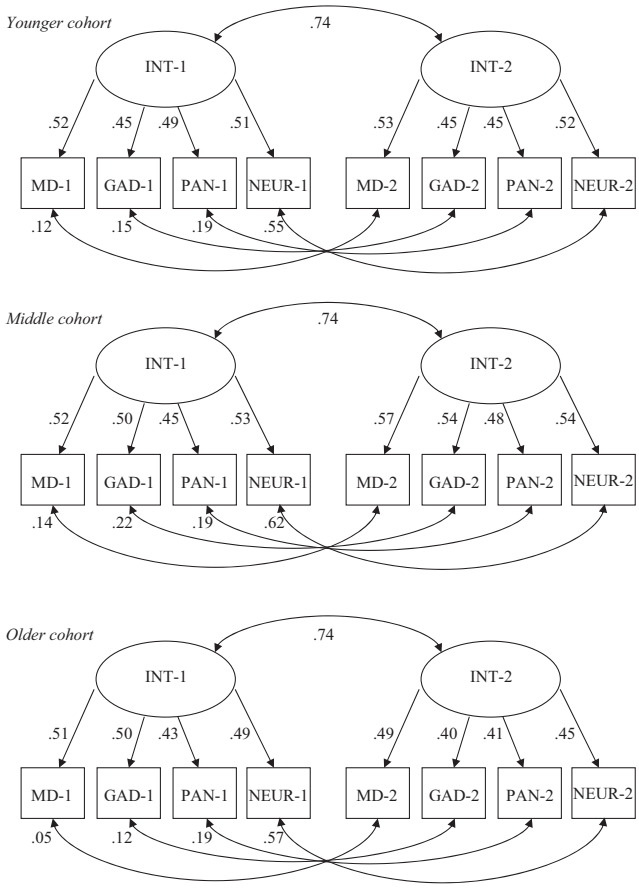

Figure 1. Standardized estimates of the best-fitting internalizing (INT) model in younger (35 years and under; $n = 1{,}729$), middle (36–50 years; $n = 2{,}719$), and older (over 50 years; $n = 2{,}601$) cohorts with equal stabilities over time. Note: MD is major depression, GAD is generalized anxiety disorder, PAN is panic disorder, NEUR is neuroticism. −1 suffix indicates MIDUS-1 data and −2 suffix indicates MIDUS-2 data. Some parameter estimates constrained to equality appear slightly different due to standardization.

invariance, because they each imposed only some of the parameter constraints of the full metric invariance model. These models examined metric invariance across groups and over time. During model evaluation, examination of each group's contributions to the overall chi-square goodness of fit test (CSGOF) indicated that the younger and middle cohorts generally contributed similarly to the CSGOF (e.g., contributions of 26.184 and 30.189, respectively, in Model 2) whereas the older cohort contributed markedly more (i.e., 51.758). This suggested the possibility of constraining younger and middle cohort parameters to equality (while freely estimating older cohort parameters); Model 4 supported this hypothesis, indicating older group parameters might be unique. Model 6, testing full metric invariance, provided borderline acceptable fit but failed to balance fit and parsimony as well as Model 4. Again, examination of group contributions to the CSGOF revealed a disproportionately large contribution from the older cohort.

Strong invariance. Strong invariance follows metric invariance with the next most stringent constrains. In addition to metric

Table 1
Model Specifications and Fit Index Values

Model	Description	RMSEA	CFI	TLI	BIC
1	Unconstrained structural model. Factor means were fixed to zero and all other parameters free to vary. INT factors, and residuals of indicator variables, were allowed to covary with themselves across time points (e.g., MD-1 residuals were correlated with MD-2 residuals).	.017	.989	.979	52,656.111
2	Metric invariance over time model. Same as Model 1, except factor loadings were constrained to equality within each group over time (e.g., MD-1 and MD-2 had the same loading in the younger cohort, which differed from loadings in the other age cohorts).	.020	.982	.973	52,636.832
3	Partial metric invariance model (over time and younger/middle cohorts). Same as Model 2, except factor loadings from younger and middle cohorts were also constrained to equality across cohorts as well as over time (e.g., MD-1 and MD-2 had the same loading, which was also equal across younger and middle cohorts).	.019	.982	.975	52,613.074
4	Partial metric invariance model (over time and younger/middle cohorts; freely estimated older cohort). Same as Model 3, except older cohort INT-1 and INT-2 factor loadings were freely estimated.	.016	.988	.982	52,590.828
5	Metric invariance across cohorts model. Same as Model 1, except INT-1 factor loadings were constrained to equality across all cohorts and INT-2 factor loadings were constrained to equality across all cohorts (i.e., INT-1 and INT-2 had different loadings).	.026	.966	.953	52,714.037
6	Full metric invariance model. Same as Model 1, except factor loadings were constrained to equality over time and across cohorts (i.e., only four loadings, one for each indicator variable, were estimated in total).	.027	.960	.948	52,751.278
7	Full strong invariance model. Same as Model 6, except (1) latent means were freely estimated over time and across cohorts, and (2) each indicator's intercept was constrained to equality over time and across cohorts.	.032	.930	.928	52,768.906
8	Hybrid metric/strong invariance model (strong in younger/middle cohorts; metric in older cohort). Same as Model 7 for younger and middle cohorts. For the older cohort, factor loadings were constrained to equality over time (and were unrelated to other cohorts' loadings).	.023	.968	.962	52,606.043
9	Partial strong invariance model (strong in younger/middle cohorts; freely estimated older cohort). Same as Model 8 above, except older cohort INT-1 and INT-2 parameters (loadings, intercepts) were freely estimated.	.022	.973	.966	52,583.798
10	Full strict invariance model. Same as Model 7, except each indicator's residual variance was constrained to equality over time and across cohorts.	.052	.778	.815	54,146.560
11	Partial strict invariance model (strict in younger/middle cohorts; freely estimated older cohort). Same as Model 10 above, except older cohort INT-1 and INT-2 parameters (loadings, intercepts, residuals) were freely estimated.	.046	.850	.856	53,655.422
12	Correlated residual equality model (in younger/middle cohorts). Same as Model 9 above, except the correlations of residuals over time were constrained to equality across younger and middle cohorts (e.g., MD-1 and MD-2 residual correlation was constrained to equality across younger/middle cohorts). Residual correlations over time in the older cohort were freely estimated.	.021	.973	.968	52,569.011

Note. INT = internalizing; MD = major depression; Suffixes −1 and −2 indicate MIDUS-1 and MIDUS-2 data, respectively. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; BIC = Bayesian information criterion.

invariance requirements, intercepts are constrained to equality. As such, all six INT factors had the same intercepts in the full strong invariance model (Model 7). Latent means were freely estimated across groups and time points. Strong invariance would indicate that the covariance structure of MD, GAD, PAN, and NEUR, and INT overall, was similar across cohorts and over time within individuals (metric invariance); further, it would indicate that group differences in the average levels of each indicator could be accounted for by group differences in the mean levels of latent INT between the cohorts. For example, in a strong invariance model, higher MD prevalence in the younger cohort than the older cohort would emerge from younger cohort members having higher average INT levels than older cohort members. Strong invariance yielded poor fit and was inferior to metric invariance by BIC. These findings, and examination of intercepts below, suggested the younger and middle cohorts might show equivalent intercepts, which we tested in Models 8 and 9. While full metric fit better than full strong invariance, Model 9 (partial strong) provided the best fit of any model tested so far.

Why did strong invariance fail to fit well? Examination of the groups' intercepts from the full metric invariance model (Model 6) suggested different average levels of endorsement of psychopathology and NEUR between the younger/middle cohorts and the older cohort. While the younger and middle cohorts' intercepts were quite similar, those of the older cohort were lower for every indicator (e.g., MD-1 intercept = .271 for younger and middle and

.153 for the older cohort). A similar pattern of declining INT means was seen. With the younger cohort's means constrained to zero, the middle cohort's INT-1 and INT-2 means were, respectively, .01 and −.26; the older cohort's means were −.36 and −.61. We tested INT mean stability within groups over time by constraining intercepts to equality across both time points in each group. With INT-1 means constrained to zero, INT-2 means decreased significantly over time in the younger ($z = -.096$; $p = .01$), middle (−.234; $p < .01$), and older cohorts (−.318; $p < .01$).

Strict invariance. Strict invariance (Model 10) is the most stringent model. It requires strong invariance as well as indicators' residual variances be constrained to equality across groups. In addition to metric and strong invariance inferences, strict invariance would indicate there were no age differences in indicator variances for which the latent INT did not account. Model 11 tested strict invariance in the younger and middle cohorts only. Both showed poor fit.

INT Stability Analyses

The study's second goal was to compare INT stability across age groups, which required use of the best fitting model (Model 9) for these between-person age-cohort comparisons. To determine whether each correlated residual path (which might affect INT stability) could be constrained to equality across groups, we tested all possible combinations of (un)constrained correlated residuals. The best fitting model (Model 12) constrained younger and middle cohorts' correlated residuals to equality while the older cohorts' were freely estimated. Model 12 fit well and had most favorable BIC value of any model. The stability coefficients (i.e., the INT-1/INT-2 correlation within each group) were $r = .79$ (younger), .72 (middle), and .74 (older). Constraining these coefficients to equality ($r = .74$) yielded acceptable fit (CFI = .973, TLI = .969, RMSEA = .021) and the most favorable BIC value (BIC = 52,553.747). A chi-square difference test between this model and Model 12 was not significant, $\chi^2[2] = 1.037$, $p = .595$; stability was thus moderate-to-high and equal across cohorts. Parameters for the final model are given in Figure 1.

Discussion

Although it is associated with important public health outcomes, little is known about how INT relates to aging. Given INT's promise for reconceptualizing disorders, framing the "metastructure" of DSM-5 and ICD-11, explaining comorbidity, and serving as a target of broad-band treatments, a richer understanding of this construct is critical. Of great importance, due to a rapidly aging population, is whether the INT structure shows within-person change over time and between-person age-cohort differences. Similarly, little is known about the long-term stability of INT at different points in adult development. The present study addressed INT structure and long-term stability in an aging context.

INT Structure Between Cohorts and Within Cohorts Over Time

Factorial invariance analyses indicated that a metric invariance model fit the data from all three age groups, indicating that INT had the same meaning across the cohorts and supporting comparisons of the characteristics of INT (e.g., variance, covariances) across groups (Gregorich, 2006). Evidence for strong invariance was found in the younger and middle cohorts, supporting stronger inferences between these two cohorts: No response biases existed between these two groups, and group-level mean differences in indicators were due to group-level mean differences in INT. The failure of full strong and strict invariance suggested that observed indicator levels and variances were not directly comparable across all cohorts. Results also supported a degree of within-cohort developmental invariance. Metric invariance indicated the meaning of INT was equivalent over nearly a decade. The partial strong invariance across MIDUS-1 and MIDUS-2 in the younger and middle cohorts indicated that within-cohort developmental changes in indicator means reflected changes in INT means; INT within the older cohort was best estimated freely, meaning its structure was not developmentally invariant.

The dissimilarity of the older cohort to the others might be due simply to cohort effects or response patterns (e.g., "nay-saying"). These differences in INT structure might also reflect age-related psychological changes. Across and within cohorts developmentally, INT means and indicator intercepts were reduced in almost every case, which is broadly congruent with previous findings that prevalence rates of mood and anxiety disorders tend to be lower in older individuals (e.g., Kessler et al., 1994; Regier et al., 1988). However, the lack of strong and strict invariance between the younger/middle and older groups suggests that between-cohort comparisons of INT means, and symptom counts and variances, are not justified; this finding likely extends to comparisons of diagnoses (dichotomized symptom counts) and thus prevalence rates. Reductions in INT means and intercepts also seem compatible with evidence of improving emotion regulation with age, suggesting that factors such as emotional control might have led to higher emotional stability, and lower levels of psychopathology, in the older cohort (Gross et al., 1997). Similarly, increased emotion differentiation, which is positively related to emotional regulation (Barrett, Gross, Christensen, & Benvenuto, 2001), is another possible mechanism. Because our findings did not support the most stringent levels of invariance, especially for the older cohort, they appear less consistent with theories of similar emotional experience across age cohorts (e.g., Lawton et al., 1992). Our results appear more in line with notions of age-related increases in emotional maturity and health (e.g., Carstensen et al., 1997; Isaacowitz et al., 2000; cf. Teachman et al., 2007), although inferences about within-person development change processes drawn from between-person age-cohort comparisons must be made with some caution.

Long-Term Internalizing Stability Within Cohorts Developmentally

This was the first study to examine long-term INT stability in the context of aging. INT showed equivalently moderate-to-high stability over nearly a decade in all three cohorts. Thus, INT structure differed somewhat across cohorts but its *stability* was similar. While disorders such as MD may manifest and remit over time, within-cohort analyses indicated the underlying liability to develop them is quite stable developmentally. This is congruent with previous research showing that primarily genetic factors drive INT (co)variance (Hettema et al., 2006; Kendler et al., 2011). In a

sense, our within-cohort invariance results can also be interpreted in terms of stability: INT structure showed a degree of stability developmentally within cohorts.

Taken together, these results support the role of INT in nosology and public health. Our findings suggest that a DSM-5 and ICD-11 INT metastructural organization for mood and anxiety disorders would be generally appropriate for adults in different age cohorts. Interventions that target INT liability (e.g., Barlow et al., 2011), rather than its specific manifestations in mood and anxiety disorders, seem justified with regard to clients in different age cohorts as well, given that INT (a) was relatively stable over time and (b) showed a notable degree of invariance across cohorts.

Limitations and Future Directions

This study had several limitations. First, we examined only INT, as there were inadequate indicators to model externalizing liability; thus, questions about the age-related structure and stability of externalizing remain unanswered. Second, congruent with previous research (e.g., South & Krueger, 2008), this study modeled four indicators of INT. Additional research would benefit from using a broader array of indicator variables and modeling the bifurcated INT model found in some data (i.e., distress and fear INT subfactors; Krueger & Markon, 2006). Third, the indicators used limit the generalizability of all latent variable studies; INT studies sometimes use different indicators, and thus our INT factor may not be completely isomorphic with INT from other studies. Fourth, we examined INT invariance rather than the invariance of the indicators; our results should not be interpreted as providing support for invariance of MD, GAD, PAN, or NEUR. This is an important issue for further research, but one that is different from our focus on INT. Finally, future studies would benefit from investigating INT structure and stability over even longer time periods, with more tightly defined cohorts, and via multiple assessments (e.g., via latent growth curve modeling).

References

Allemand, M., Zimprich, D., & Hertzog, C. (2007). Cross-sectional age differences and longitudinal age changes of personality in middle adulthood and old age. *Journal of Personality, 75,* 323–358. doi:10.1111/j.1467-6494.2006.00441.x

Andrews, G., Goldberg, D. P., Krueger, R. F., Carpenter, W. T., Jr., Hyman, S. E., Sachdev, P., & Pine, D. S. (2009). Exploring the feasibility of a meta-structure for DSM-V and ICD-11: Could it improve utility and validity? *Psychological Medicine, 39,* 1993–2000. doi:10.1017/S0033291709990250

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. (2011). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide.* New York: Oxford University Press.

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion, 15,* 713–724. doi:10.1080/02699930143000239

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

Carstensen, L. L., Gross, J. J., & Fung, H. H. (1997). The social context of emotional experience. In K. W. Schaie & M. P. Lawton (Eds.), *Annual review of gerontology and geriatrics* (Vol. 17, pp. 325–354). New York: Springer Publishing.

Eaton, N. R., South, S. C., & Krueger, R. F. (2010). The meaning of comorbidity among common mental disorders. In T. Millon, R. Krueger, & E. Simonsen (Eds.), Contemporary directions in psychopathology: Scientific foundations of the *DSM-V* and *ICD-11* (2nd ed., pp. 223–241). New York: Guilford Press Publications.

Fergusson, D. M., Horwood, L. J., & Boden, J. M. (2006). Structure of internalizing symptoms in early adulthood. *The British Journal of Psychiatry, 189,* 540–546. doi:10.1192/bjp.bp.106.022384

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44,* S78–S94. doi:10.1097/01.mlr.0000245454.12228.8f

Griffith, J. W., Zinbarg, R. E., Craske, M. G., Mineka, S., Rose, R. D., Waters, A. M., & Sutton, J. M. (2010). Neuroticism as a common dimension in the internalizing disorders. *Psychological Medicine, 40,* 1125–1136. doi:10.1017/S0033291709991449

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam Skorpen, C., & Hsu, A. Y. C. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging, 12,* 590–599. doi:10.1037/0882-7974.12.4.590

Hettema, J. M., Neale, M. C., Myers, J. M., Prescott, C. A., & Kendler, K. S. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. *American Journal of Psychiatry, 163,* 857–864. doi:10.1176/appi.ajp.163.5.857

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

Isaacowitz, D. M., Charles, S. T., & Carstensen, L. L. (2000). Emotion and cognition. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 593–632). Mahwah, NJ: Erlbaum.

Jeste, D. V., Alexopoulos, G. S., Bartels, S. J., Cummings, J. L., Gallo, J. J., Gottlieb, G. L., & Lebowitz, B. D. (1999). Consensus statement on the upcoming crisis in geriatric mental health. *Archives of General Psychiatry, 56,* 848–853. doi:10.1001/archpsyc.56.9.848

Kendler, K. S., Aggen, S. H., Knudsen, G. P., Røysamb, E., Neale, M. C., & Reichborn-Kjennerud, T. (2011). The structure of genetic and environmental risk factors for syndromal and subsyndromal common *DSM–IV* axis I and all axis II disorders. *American Journal of Psychiatry, 168,* 29–39. doi:10.1176/appi.ajp.2010.10030340

Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H. U. (1998). The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). *International Journal of Methods in Psychiatric Research, 7,* 171–185. doi:10.1002/mpr.47

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., . . . Kendler, K. S. (1994). Lifetime and 12-month prevalence of *DSM–III–R* psychiatric disorders in the United States. *Archives of General Psychiatry, 51,* 8–19.

Kessler, R. C., Ormel, J., Petukhova, M., McLaughlin, K. A., Green, J. G., Russo, L. J., . . . Üstün, T. B. (2011). Development of lifetime comorbidity in the World Health Organization World Mental Health Surveys. *Archives of General Psychiatry, 68,* 90–100. doi:10.1001/archgenpsychiatry.2010.180

Kramer, M. D., Krueger, R. F., & Hicks, B. M. (2008). Internalizing and externalizing liability factors account for gender differences in the prevalence of common psychopathological syndromes. *Psychological Medicine, 38,* 51–61. doi:10.1017/S0033291707001572

Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (*DSM–III–R*): A longitudinal-epidemiological study. *Journal of Abnormal Psychology, 108,* 216–227. doi:10.1037/0021-843X.107.2.216

Krueger, R. F., Chentsova-Dutton, Y. E., Markon, K. E., Goldberg, D., &

Ormel, J. (2003). A cross-cultural study of the structure of comorbidity among common psychopathological syndromes in the general health care setting. *Journal of Abnormal Psychology, 112,* 437–447. doi:10.1037/0021-843X.112.3.437

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2,* 111–133. doi:10.1146/annurev.clinpsy.2.022305.095213

Lahey, B. B. (2009). Public Health significance of neuroticism. *American Psychologist, 64,* 241–256. doi:10.1037/a0015309

Lawton, M. P., Kleban, M. H., Rajagopal, D., & Dean, J. (1992). Dimensions of affective experience in three age groups. *Psychology and Aging, 7,* 171–184. doi:10.1037/0882-7974.7.2.171

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525–543. doi:10.1007/BF02294825

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (4th ed.). Los Angeles: Author.

Radler, B. T., & Ryff, C. D. (2010). Who participates? Accounting for longitudinal retention in the MIDUS National Study of Health and Well-Being. *Journal of Aging and Health, 22,* 307–331. doi:10.1177/0898264309358617

Regier, D. A., Boyd, J. H., Burke, J. D., Rae, D. S., Myers, J. K., Kramer, M., . . . Locke, B. Z. (1988). One-month prevalence of mental disorders in the United States: Based on five epidemiologic catchment area sites. *Archives of General Psychiatry, 45,* 977–986.

Regier, D. A., Narrow, W. E., Kuhl, E. A., & Kupfer, D. J. (Eds.). (2011). *The conceptual evolution of DSM-5.* Arlington, VA: American Psychiatric Publishing, Inc.

South, S. C., & Krueger, R. F. (2008). Marital quality moderates genetic and environmental influences on the internalizing spectrum. *Journal of Abnormal Psychology, 117,* 826–837. doi:10.1037/a0013499

Teachman, B. A., Siedlecki, K. L., & Magee, J. C. (2007). Aging and symptoms of anxiety and depression: Structural invariance of the tripartite model. *Psychology and Aging, 22,* 160–170. doi:10.1037/0882-7974.22.1.160

Vollebergh, W. A. M., Iedema, J., Bijl, R. V., de Graaf, R., Smit, F., & Ormel, J. (2001). The structure and stability of common mental disorders. *Archives of General Psychiatry, 58,* 597–603. doi:10.1001/archpsyc.58.6.597

Wang, P. S., Berglund, P., & Kessler, R. C. (2000). Recent care of common mental disorders in the United States: Prevalence and conformance with evidence-based recommendations. *Journal of General Internal Medicine, 15,* 284–292. doi:10.1046/j.1525-1497.2000.9908044.x

Watson, D. (2009). Differentiating the mood and anxiety disorders: A quadripartite model. *Annual Review of Clinical Psychology, 5,* 221–247. doi:10.1146/annurev.clinpsy.032408.153510

Received May 5, 2010
Revision received April 21, 2011
Accepted April 22, 2011 ■