Joormann, J., & Stöber, J. (1997). Measuring facets of worry: A LISREL analysis of the Worry Domains Questionnaire. *Personality and Individual Differences*, *23*(5), 827-837.

Measuring Facets of Worry: A LISREL Analysis of the Worry Domains Questionnaire

Jutta Joormann* and Joachim Stöber†

*Department of Research Methods, Social Psychology, and Developmental Psychology,

†Department of General Psychology, Biological Psychology, and Cognitive Psychology,

Free University of Berlin, FB 12, Habelschwerdter Allee 45, D-14195 Berlin, Germany

Summary

In the development of the Worry Domains Questionnaire (WDQ; Tallis, Eysenck, & Mathews, 1992) for the measurement of nonpathological worry, Tallis et al. had used cluster analytical procedures to establish the number of worry domains. The resulting structure of the WDQ, however, was never adequately tested. This study therefore examined the WDQ's structure by use of confirmatory factor analysis comparing models of different factor structures. In a first sample of 466 participants, a five-factor model yielded the best fit to the data, characterized by highly correlated yet distinct domains of everyday worrying as they were originally proposed. This model was cross-validated with a second sample of 503 participants, showing stable factor loadings across samples. Whereas these analyses displayed a good fit of the five-factor representation for the item-based models, overall fit of all models was more prominent when items were aggregated (subscale models). Implications of the results and suggestions for future research are discussed.

Keywords

anxiety / questionnaires / structural equation modeling / factor structure / factorial validity

Measuring Facets of Worry: A LISREL Analysis of the Worry Domains Questionnaire

Introduction

For over three decades, worry has been a fruitful concept in anxiety research. The starting point was the seminal work of Liebert and Morris (1967) who first discriminated between a cognitive component (worry) and a physiological component (emotionality) in test anxiety and then demonstrated that worry, and not emotionality, was responsible for the detrimental effects that test anxiety had on performance (Morris & Liebert, 1970). While these findings stimulated a great deal of worry research for the consecutive two decades, the primary focus of this research remained on academic performance and other achievement-related settings (cf. the meta-analysis of Seipp, 1991).

At the beginning of the 80s, a second line of worry research emerged that broadened this rather narrow focus. Excessive uncontrollable worry about a range of different topics was recognized as the central feature of generalized anxiety disorder (GAD) in the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R; American Psychiatric Association, 1987). Although this second line of research centered mainly on pathological worrying, some researchers pointed out that attention should also be paid to worry outside of the clinical context (e.g., Stöber, 1996; Tallis, Eysenck, & Mathews, 1992). In particular, Davey (1994) claimed that the worry experienced by GAD patients was only the extreme end of a continuum ranging from abnormal, pathological worrying to normal, nonpathological worrying. Worry is actually a frequent phenomenon in the everyday lives of normal, well-adjusted people, and this type of worry is sometimes even conceived of as a constructive activity that helps to analyze problems and motivates problem-solving. More often, however, the costs outweigh the benefits, as worry produces emotional discomfort, exaggerates problems, and leads to a pessimistic view of things to come (Tallis, Davey, & Capuzzo, 1994).

With respect to the development of worry questionnaires, two approaches have emerged: "content-free" measures and "content-based" measures. Coming mainly from research on pathological worry, content-free measures are instruments that assess the excessiveness, duration, and uncontrollability of worry and associated stress. Without asking for specific contents, these questionnaires measure how typical in general the symptoms of worry are for the individual. Examples of this approach are the Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990) and the subscale "meta-worry" of the Anxious Thoughts Inventory (AnTI; Wells, 1994). In contrast, the content-based approach comes mainly from research in nonpathological worry. Respondents give intensity or frequency ratings with respect to a list of worry topics. Examples of this approach are the subscales "social worry" and "health worry" of the AnTI (Wells, 1994), the Worry Scale for the elderly (Wisocki, Handen, & Morse, 1986), the Student Worry Scale (Davey, Hampton, Farrell, & Davidson, 1992), and the Worry Domains Questionnaire (Tallis et al., 1992).

For the comprehensive study of individual differences in the level of nonpathological worry, Tallis et al.'s Worry Domains Questionnaire (WDQ) seems the most promising instrument up to date. Firstly, unlike the content-free measures that resulted from clinical experience with GAD patients and thus focus on the pathological, dysfunctional aspects of worry, the WDQ was developed as a general measure of nonpathological worry for nonclinical adult samples. The initial item pool for the WDQ was collected from a nonclinical sample (cf. Method section for details), and the questionnaire covers a broad range of everyday worries, not only social concerns and health concerns. Moreover, the WDQ is applicable to a wide range of different populations, not only to students or to elderly people. Finally, with Cronbach's αs above .90 (Davey, 1993; Stöber, 1995), the WDQ provides total scores with an internal consistency well above the .80 level that Carmines and Zeller (1979, p. 51) recommended for widely-used scales. Therefore, it is not surprising that the WDQ was quickly adopted internationally by various research teams investigating individual differences in worry (e.g., Dugas, Freeston, Doucet, Lachance, & Ladouceur, 1995; East & Watts, 1994; Schwarzer, 1993; Stöber, 1996).

Aim of the Present Study

Despite the increasing number of studies using the WDQ, some important questions remain. In particular, the factor structure of the WDQ has not yet been satisfactorily tested. Tallis et al. (1992) claimed that the WDQ covered five different domains of everyday worry (Relationships, Lack of Confidence, Aimless Future, Work Incompetence, and Financial). These subscales were formed by selecting items most representative of the worry domains that were found using cluster analytical procedures, followed by expert ratings for their coherence. The structure of the questionnaire that was a result of this procedure, however, was never adequately tested.

With a sample of 224 participants, Stöber (1995) conducted a first exploratory factor analysis on the WDQ items which found only four substantive factors. Moreover, these factors did not correspond well to the domains suggested by Tallis et al.: Whereas all of the Work Incompetence items and all of the Financial items displayed substantial loadings only on their respective factors, most of the items from the other three subscales showed loadings on both of the two remaining factors (see Stöber, 1995, Table 3). Although all five-item subscales had high internal consistencies (Cronbach's αs from .75 to .86), at the same time several of them were highly intercorrelated. Relationships, Lack of Confidence, and Aimless Future showed correlations between .66 and .71 whereas Financial correlated only .18 to .25 with the other WDQ subscales.

Because exclusive reliance on exploratory data analysis techniques so far has lead only to inconclusive results, the purpose of the present study was to test the factor structure of the WDQ by use of confirmatory factor analysis (Jöreskog & Sörbom, 1989). This approach permits a more explicit test of the postulated five-factor structure and allows for straightforward comparison of alternative models. It also provides a more formal and convenient way to test hypotheses concerning factor structure and parameters across different samples, thereby allowing for a cross-validation of the obtained results.

Method

Participants

The data for the subsequent analyses were taken from various studies conducted by the second author. The first sample, coming from studies conducted from December 1994 to January 1996, consisted of 466 participants (270 female, 196 male), of which 194 were psychology students at the Free University of Berlin. The mean age in this sample was 27.92 years (SD = 7.07). The second sample, coming from studies conducted from March 1996 to June 1996, consisted of 503 participants (315 female, 188 male), of which 136 were psychology students. The mean age of the participants in the second sample was 27.32 years (SD = 7.43). In all studies, participants had volunteered to take part, and psychology students had received course credits for participation.

The Worry Domains Questionnaire

The Worry Domains Questionnaire (WDQ) was developed by Tallis et al. (1992) for the measurement of nonpathological worry. First, Tallis et al. distributed one hundred questionnaires to collect a large pool of various worry items. Seventy-one questionnaires were returned. The responses were used to construct a 155-item General Worry Questionnaire. Then, a second sample of 95 participants rated each worry item for both frequency and intensity. Following theoretical suppositions by Eysenck (1984) who postulated the existence of organized clusters of worry-related information in long term memory, the frequency and intensity ratings were analyzed using cluster analysis.

This procedure, however, had some problematic aspects: Although cluster analysis resulted in semantically cohesive clusters of between four and seven items, clusters were selected using an aggregation rule that Tallis et al. (1992) themselves described as an "arbitrary method" (p. 163). Furthermore, frequency ratings and intensity ratings produced different numbers of clusters. Therefore, both sets of clusters were presented to ten judges who rated them "for coherence . . . in terms of an underlying theme" (p. 164). This procedure resulted in a formation of six clusters in both sets. Consequently, frequency and intensity ratings were collapsed, and from each of the six content domains, the five items that were most commonly endorsed in both ratings were selected for inclusion in the WDQ. The pilot version of the WDQ therefore contained six domains. However, the sixth domain, Socio-Political, was consecutively omitted because of low correlations with the other domains and endorsement rates highly influenced by social desirability (cf. Tallis et al., 1992; Tallis, Davey, & Bond, 1994).[*]

As presented by Tallis et al. (1994), the final version of the WDQ now covers only five domains of nonpathological worry. These are labeled (abbreviations in brackets): Relationships

[*]In a first application of this questionnaire, Tallis, Eysenck, and Mathews (1991) presented yet another version of the WDQ which contained "Physical Threat" as a sixth domain. This domain, however, was also omitted from the final version.

(Rel), Lack of Confidence (L of C), Aimless Future (Aim Fut), Work Incompetence (Work Inc), and Financial (Fin). With five items per domain, the WDQ comprises 25 items. Items are ordered randomly. Compared to scale-wise item blocking, this approach has been shown to improve reliability of the questionnaire subscales (Krampen, 1993). The instructions of the WDQ ask respondents to tick one of five boxes after each item, reflecting their amount of worry. These are labeled: "Not at all" (scoring 0), "A little" (1), "Moderately" (2), "Quite a bit" (3), and "Extremely" (4). The English version of the WDQ is shown in the Appendix. In the present study, the German translation was used (Stöber, 1995, p. 56).

The Confirmatory Factor Analysis Approach

To investigate the factor structure of the WDQ, confirmatory factor analysis was employed to evaluate models of different factor structures and compare them with respect to their fit to the data. This analysis comprised three steps: First, an effort was made to test the adequacy of the five-factor model postulated by Tallis et al. (1992). Secondly, a sequence of alternative models of different dimensionalities was inferred and subsequently compared with respect to model fit. Finally, analyses were conducted using data from the second independent sample to test the invariance of the models' parameter estimates across the two samples.

Two different sets of models were evaluated: item-based models and subscale models. For item-based models, the responses to the 25 WDQ items were used as indicators of the postulated latent variables (i.e., factors in common factor analysis). Thus, each domain of the WDQ was represented as a latent variable with the five domain-items all loading individually on their respective factor. However, the use of single items as indicators in factor analysis has been criticized. Bernstein and Teng (1989), for example, found evidence for multi-dimensionality when factoring single items instead of multi-item scales. Moreover, the overall fit of models using single items could be underestimated, when compared to models using subscales, merely because of the lower reliability of single items and the greater number of estimated parameters. Therefore, the present study also investigated subscale models in the confirmatory factor analysis of the WDQ. Following a standard procedure (see, e.g., Bagozzi & Foxall, 1995), the five items indicative of a factor were randomly split into subgroups of two and three items, thus creating multi-item subscales to serve as indicators.

All analyses were conducted with correlation and covariance matrices. Factor loadings, factor intercorrelations, and uniquenesses were set free to be estimated with LISREL 7.2 (Jöreskog & Sörbom, 1989). As a consequence of the estimation procedure, factor intercorrelations were automatically corrected for attenuation due to the unreliability of measures (cf. Jöreskog & Sörbom, 1989).

Starting with the structure proposed by Tallis et al. (1992), the initial model tested was the five-factor model. To account for item covariations, this model postulated five intercorrelated but distinct factors that corresponded to the five worry domains described above. The different magnitudes of the domain intercorrelations in this model were then used to inspect alternative models with different factor structures: First, the five-factor model was compared to a three-factor model that integrated the highly intercorrelated factors Relationships, Lack of Confidence, and Aimless Future into one factor and left Work Incompetence and Financial as two separate factors. The next alternative model implied by the domain intercorrelations was a two-factor model in which Relationships, Lack of Confidence, Aimless Future, and Work Incompetence were integrated into one factor and only Financial was left as a separate factor. As the third alternative model, a one-factor model was considered in which all WDQ items formed one single factor. These different models were then compared in terms of both fit to the data and parsimony of representation.

Assessment of Fit and Model Comparison

One common measure to assess the fit of a model is the likelihood-ratio χ^2 statistic. This statistic is used to test the null hypothesis that a specific model reproduces the population covariance-matrix of the observed variables. A well-fitting model is associated with a nonsignificant χ^2 statistic, indicating that any discrepancy between the observed matrix and the matrix generated by the theoretical model is not significant. However, when assessing the fit of LISREL models, it is widely advocated not to rely only on one single fit index, but instead to compare different fit indices. Particularly, it is recommended not to rely only on the χ^2 test because of its dependence on sample size (Bentler, 1990; Bentler & Bonett, 1980). In large samples, already trivial deviations of a hypothesized model from the true model may lead to a rejection of the model, whereas in small samples, even large deviations of a hypothesized model from the true model may go undetected. Moreover, the χ^2 test does not provide a direct estimate of the degree of model fit because it is not normed from zero to unity.

An additional approach to the assessment of model fit is the use of incremental fit indices (also called relative fit indices). These indices are based on the comparison of the fit of a hypothesized model with the fit of a baseline model, such as the so-called "null model". The null model is a no-factor model in which all variables are assumed to be uncorrelated (i.e. only error variances are estimated). These fit indices are called incremental fit indices because a hypothesized model is compared to a more restricted, nested model. Especially for the objective of the present study (the investigation whether a particular model, i.e., the five-factor model, would provide a better fit than alternative, more restricted models), incremental fit indices provide a major means to test model fit.

A well-known incremental fit index is the relative noncentrality index (RNI; McDonald & Marsh, 1990). This index provides an unbiased estimate of its corresponding population value, and thus should be independent of sample size. Ranging from zero to unity, the RNI is a normed-fit index and can be thought of as a measure of how much variation is accounted for by a given model. For well-fitting models, values should be greater than or equal to .90, whereas values less than .90 indicate that a significant amount of variation remains to be explained (Bentler & Bonett, 1980). Monte Carlo studies have shown that the RNI performed well for sample sizes from 50 to 1,600 and produced unbiased estimates that are low in variability (Bentler, 1990).

However, one problematic characteristic of normed-fit indices is that the fit of a model can be increased simply by freeing up parameters to be estimated (James, Mulaik, & Brett, 1982). Each additional parameter that is freed removes one constraint on the final solution, with the result that the reproduced data matrix will better fit the sample data matrix. Thus, a two-factor model may fit the data better than the single factor model simply because there is one additional parameter being estimated. To compensate for this problem, we also calculated the parsimonious fit index (PNFI), a measure that explicitly takes model parsimony into account (Mulaik, James, Van Alstine, Bennett, Lind, & Stilwell, 1989).

To assess the fit of the various models, we followed the general guidelines proposed by Marsh (1990): (1) Solutions were inspected to decide if they were well defined. In particular, convergence of iterative procedures to a proper solution, permissible ranges of parameter estimates, and values of standard errors were considered. (2) Parameter estimates were examined in relation to the a priori model and common sense. (3) Fit indices and χ^2 tests were evaluated and compared to alternative models where appropriate.

Additionally, cross-validation procedures were employed to assess the goodness of fit of a particular model by testing whether the factor structure and the parameter estimates were identical across the two independent samples. This was done by fixing the key parameters of the five-factor model to be equal across samples and then by comparing the fit of this five-factor model to the data of the second sample with the fit of the model where no such equality constraints were made. The invariance of the factor loadings was tested by investigating whether the measures in both samples indicated the same factors and the same factor structure. Additionally, a test of the equality of correlations among the factors across samples was performed to examine if the model could be replicated (and if not, which parts of the model were difficult to replicate). These analyses were performed for the five-factor model both on the item-based level and the subscale level.

Results

Test of the Five-Factor Model

With respect to the parameter estimates and the overall goodness-of-fit measures, the analyses based on correlations and the ones based on covariances did not produce any significantly different results. Therefore, only the results of the correlational analyses are reported here. For the five-factor model using single items, Table 1 shows the standardized factor loadings, uniquenesses, and standard errors of the parameter estimates for the two samples. Factor loadings were moderate to high, ranging from .41 to .84 in the first sample and from .34 to .85 in the second sample, with standard errors generally being of small magnitude.

Table 2 displays the descriptive statistics, Cronbach's αs, and intercorrelations of the item composites indicative of the five respective domains. Whereas Relationships, Aimless Future, and Work Incompetence showed internal consistencies in the .70s, Lack of Confidence and Financial had Cronbach's αs in the .80s. Corroborating the analyses by Davey (1993) and Stöber (1995), the WDQ total score showed an internal consistency of .90 in both samples. Regarding the intercorrelations among the five postulated factors in the two samples, the domains of Relationships, Lack of Confidence, and Aimless Future were highly correlated among each other (.71 to .79) and showed smaller correlations with Work Incompetence (.46 to .72) and only moderate correlations with Financial (.22 to .44). Whereas this correlational pattern was in accord with the pattern found by Stöber (1995), the correlations were altogether of higher magnitude, an effect due to the correction for attenuation inherent in the LISREL approach.

Model Comparison

Item-based models. The fit indices for the evaluations and model comparisons of the item-based models are listed in Table 3. On the basis of the χ^2 tests, all models had to be rejected although, as mentioned previously, this was probably due to the large sample sizes. The values of the relative noncentrality index (RNI) were also below .90 for all models, suggesting that a significant amount of variation remained to be explained. Regarding the difference of the models, all χ^2 difference tests were significant, indicating that the models provided significantly different degrees of model fit. With RNI values of .86 in both samples, the best fit was achieved for the five-factor model. This was the case even, when model parsimony was taken into account. The five-factor model also yielded the best fit with respect to the parsimonious normed fit index (PNFI). However, inspection of the PNFI values reveals that the differences between the five-factor model and the three-factor model were rather small. The superiority of fit of the five-factor model only marginally compensated for the parsimony of representation of the three-factor model.

Subscale models. Table 4 displays the fit indices for all subscale models and subsequent model comparisons. As expected, overall model fit improved when two subscales (instead of five items) were used in the calculations. Whereas all the χ^2 tests were again significant for all models, RNI values of .97 and .98 (in the first sample and in the second sample, respectively) indicated an excellent fit for the five-factor model to the data. With RNI values of .93 and .92, the overall fit of the three-factor model was smaller, but also satisfactory. Model comparisons based on χ^2 difference tests indicate that the models still do provide different degrees of model fit. Considering RNI values, the best fit was again achieved for the five-factor model. As expected, the results showed that the fit improved when more factors were proposed. Only the PNFI values made an exception; taking parsimony of representation into account, the best fitting models were the three-factor model and the two-factor model. For the subscale models, the differences in model fit was not sufficiently prominent in favor of the five-factor model to compensate for loss in parsimony of representation.

Cross Validation

Item-based models. The inspection of Tables 3 and 4 already suggested that the indices of overall fit were approximately in the same order of magnitude for both samples. To further evaluate the parameter estimates of the five-factor model achieved in the first sample, the invariance of the estimates across samples was investigated by comparing the fit of the three different models to the data of the second sample: The first model, Model 1, was the baseline model for comparison in which no constraints were placed on the five-factor model to fit the data in the second sample. Model 2 imposed the constraint of equality of factor loadings across samples: the factor loadings of the five-factor model were fixed to the values that were estimated in the first sample. In Model 3, an additional constraint of equality of factor intercorrelations across samples was added. The fit of the different models was then assessed and compared by χ^2 difference tests. Following the above sequence of model comparisons, one could determine whether the total model or just parts of it could be replicated. As can be seen in Table 5, all of the tests were significant, suggesting that the estimates of factor loadings and of factor intercorrelations differed significantly, even though the overall fit of the model was approximately the same in both samples.

Subscale models. The sequential cross-validation procedure described above was repeated for the subscale models (cf. Table 6). In contrast to the item-based model, the χ^2 test for the difference between Model 1 (baseline) and Model 2 (equal factor loadings) was not significant ($p > .25$), indicating that the estimates for the factor loadings did not differ significantly between the two samples and that the two models yielded a similar fit to the data. Thus, the factor loadings estimated for the first sample were replicated in the second, independent sample. However, this did not hold for the comparison of Model 2 with Model 3 (equal factor intercorrelations). A significant χ^2 test indicated that the factor correlations differed across samples and that the estimates from the first sample could not be replicated in the second sample.

Finally, in addition to the analyses reported above, exploratory analyses were conducted to investigate potential differences of model fit with respect to the participants' gender, degree of worry, and major subject. For these variables, the overall fit of the five-factor model as well as comparisons of models of different factor structure were evaluated. First, all of the above analyses were conducted separately for male and female participants. Secondly, samples were divided by median split into high worriers and low worriers. The third variable investigated was whether the participants were psychology students or not. The results from these various subgroup analyses did not differ from the results of the total samples, indicating that these variables did not moderate the superior fit of the five-factor model.

Discussion

The findings provide strong support for a five-factor representation of the worries presented in the Worry Domains Questionnaire (WDQ). In accordance with the domains proposed by Tallis et al. (1992; Tallis, Davey, & Bond, 1994), a model with five intercorrelated yet distinct factors was found to best account for the data. These factors corresponded to the original formulations of the five worry domains labeled Relationships, Lack of Confidence, Aimless Future, Work Incompetence, and Financial. The superior fit of this five-factor model was indicated by the χ^2 tests as well as by the fit indices that involved a direct comparison of the different factor models with reference to the null model. In particular, the single-factor model and the two-factor model failed to capture the structure of the WDQ. Also when considering models that used subscales (instead of items) as indicators, the overall fit assessed by the relative noncentrality index (RNI) suggested that the five-factor representation provided an excellent account of the subscale intercorrelations.

Cross-validation with a second, independent sample corroborated the above results. Overall, only minor differences in parameter estimates for the five-factor model across the two independent samples were found. When using χ^2 difference tests, this even replicated the exact values of the factor loadings estimated in the first sample. Moreover, subgroup comparisons indicated that the fit of the five-factor model was independent of the participants' gender, degree of worry, and major subject, supporting the generality of the five-factor representation.

There were, however, some issues that deserve further scrutiny. First, when parsimony of representation, as indicated by the parsimonious normed fit index (PNFI), was taken as the standard for model comparison, the results were not as clear-cut, because the item-based models and the subscale models produced different results. Whereas the five-factor model and the three-factor model provided the best fit-parsimony ratio when single items were used as indicators, the PNFI values were best for the three-factor model and the two-factor model when subscales were used. Furthermore, although the overall model fit improved for all subscale models as compared to their respective item-based models, the PNFI values decreased.

This difference between item-based models and subscale models might not only reflect the difference in parsimony of representation but may relate to the way in which model parsimony is assessed by the PNFI. In this index, the differences of degrees of freedom of for example the five- and the three-factor model are related to the degrees of freedom of the null model. To put it differently, the parsimony of a given representation is compared to the parsimony of the baseline model. For the item-based models with their many parameters to be estimated, the reduction of parameters is only small when comparing the five-factor model and, for example, the three-factor model to the baseline model. In contrast, the subscale models have much fewer parameters to be estimated. Therefore, the difference between the five-factor model and the three-factor model is much greater in terms of degrees of freedom relative to the baseline model. Despite these caveats, the PNFI values indicated that the five-factor model and the three-factor model did not differ significantly with respect to model fit and parsimony of representation. For the subscale models, this was the more problematic, because the three-factor model yielded also a satisfactory overall fit.

The results of the cross-validation also deserve some critical consideration. According to the χ^2 difference tests, we were not able to replicate the factor intercorrelations that were estimated in the first sample in the data of the second sample. A closer inspection of Table 2 suggests that the different correlations of the Relationships and Financial factors (.23 in the first and .42 in the second sample) were responsible for that failure. Looking at previous studies with the WDQ, one finds considerable variability in the correlations among the WDQ subscales from sample to sample. With respect to Relationships and Financial, for instance, correlations have been reported to range from .41 for a working-adults sample to .30 for a student sample (Tallis, Davey, & Bond, 1994) to .25 for a combined sample (Stöber, 1995). Whereas our confirmatory factor analyses could clarify the relation of items to domains and supported the hypothesized five-domain representation of the WDQ, the degree of interrelatedness between the five domains remains unclear. Our analysis could only add to previous findings of instability of intercorrelations in the WDQ's subscales. Further research is needed to isolate the variables related to the reported differences in the subscales' intercorrelations.

Although the internal structural of the WDQ was the main focus of the present study, our results confirmed again that the WDQ provides total scores and subscale scores of high internal consistency. With respect to the psychometric properties of the WDQ, however, further studies are needed. Particularly the data base for the instrument's stability and, above all, for its validity is still small. With regard to stability, Tallis, Davey, and Bond (1994) reported test-retest correlations of .46 to .86 for the five WDQ subscales and of .79 for the WDQ scale across an interval of two to four weeks. However, since these estimates come from a sample of only 16 participants, the confidence interval around these correlations would be quite large, leaving these findings of only limited value. Therefore, Stöber (1997) conducted a study with 148 participants using a design that also included a test-retest application of the WDQ. In addition, he collected three peer ratings for each participant to estimate the validity of the participant's self-report (cf. McCrae, 1994). For his larger sample, Stöber obtained test-retest correlations of .71 to .86 for the subscales and .85 for the total score across an interval of three to four weeks. Furthermore, the WDQ total score displayed a correlation of .49 with aggregated peer ratings, indicating a convergent validity of a magnitude that was comparable to those obtained for widely-used personality measures (cf. Borkenau & Ostendorf, 1992; McCrae & Costa, 1987).

Clearly, the WDQ can be used to measure individual differences in worry and distinguish high and low worriers in samples drawn from a nonclinical adult population. The subscales might be used to assess worry in different domains, whereas the total WDQ score might provide information of general worry across domains. In this respect, however, two related questions remain: How different are the WDQ domains, and how general is the WDQ? In a previous study on the structure of worry, Eysenck and van Berkum (1992) investigated a sample of 109 worries, aggregated them into 10 worry scales, and then subjected these scales to a principal component analysis. Two principal factors emerged. The first factor (which accounted for 52.3% of the variance) subsumed worries about general social evaluation, personal fulfillment, personal relationships, and finances. The second factor (which accounted for 11.4% of the variance) subsumed worries about the physical health of close ones, social and environmental issues, nuclear and international issues, and physical health. Consequently, Factor 1 was labeled "Social Evaluation" and Factor 2 "Physical Threat".

Therefore, when adopting this broader perspective of nonpathological worry, the five factors of the WDQ do appear to represent merely the "facets" of the first factor that was found by Eysenck and van Berkum (1992). The reason for this is that, in the course of the WDQ's development, worry domains from second general worry factor "Physical Threat" were dropped (i.e., the WDQ subscales Socio-Political and Physical Threat). As a result, the WDQ now is a short, economical worry measure with high internal consistency, but a measure only of the socio-evaluative dimension of worry. Although the five domains of the WDQ comprise the major domains of worry, future research might consider including, or again re-including, the domains from the physical-health factor to construct content-based worry measures of nonpathological worry that are truly multidimensional.

References

American Psychiatric Association (Ed.). (1987). Diagnostic and statistical manual of mental disorders. Third edition--revised (DSM-III-R). Washington, DC: Author.

Bagozzi, R. P., & Foxall, G. R. (1995). Construct validity and generalizability of the Kirton Adaption-Innovation Inventory. European Journal of Personality, 9, 185-206.

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

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. Psychological Bulletin, 105, 467-477.

Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. European Journal of Personality, 6, 199-214.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills, CA: Sage.

Davey, G. C. L. (1993). A comparison of three worry questionnaires. Behaviour Research and Therapy, 31, 51-56.

Davey, G. C. L. (1994). Pathological worrying as exacerbated problem-solving. In C. G. L. Davey & F. Tallis (Eds.), Worrying. Perspectives on theory, assessment, and treatment (pp. 35-59). New York: Wiley.

Davey, G. C. L., Hampton, J., Farrell, J., & Davidson, S. (1992). Some characteristics of worrying: Evidence for worrying and anxiety as separate constructs. Personality and Individual Differences, 13, 133-147.

Dugas, M. J., Freeston, M. H., Doucet, C., Lachance, S., & Ladouceur, R. (1995). Structured versus free-recall measures: Effect on report of worry themes. Personality and Individual Differences, 18, 355-361.

East, M. P., & Watts, F. N. (1994). Worry and the suppression of imagery. Behaviour Research and Therapy, 32, 851-855.

Eysenck, M. W. (1984). Anxiety and the worry process. Bulletin of the Psychonomic Society, 22, 545-548.

Eysenck, M. W., & van Berkum, J. (1992). Trait anxiety, defensiveness, and the structure of worry. Personality and Individual Differences, 13, 1285-1290.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). Causal Analysis: Assumptions, models, and data. Beverly Hills, CA: Sage.

Jöreskog, K. G., & Sörbom, D. G. (1989). LISREL 7: A guide to the program and applications. Chicago, IL: SPSS.

Krampen, G. (1993). Effekte von Bewerbungsinstruktionen und Subskalenextraktion in der Fragebogendiagnostik [The effects of social-desirability-provoking instructions and item blocking on results of personality questionnaires]. Diagnostica, 39, 97-108.

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. Psychological Reports, 20, 975-978.

Marsh, H. W. (1990). Confirmatory factor analysis of multitrait-multimethod data: The construct validation of multidimensional self-concept responses. Journal of Personality, 58, 661-692.

McCrae, R. R. (1994). The counterpoint of personality assessment: Self-reports and observer ratings. Assessment, 1, 159-172.

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. Journal of Personality and Social Psychology, 52, 81-90.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. Psychological Bulletin, 107, 247-255.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. Behaviour Research and Therapy, 28, 487-495.

Morris, L. W., & Liebert, R. M. (1970). Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. Journal of Consulting and Clinical Psychology, 35, 332-337.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). An evaluation of goodness of fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Schwarzer, R. (1993). Streß, Angst und Handlungsregulation [Stress, anxiety, and action control] (3rd ed.). Stuttgart, Germany: Kohlhammer.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. Anxiety Research, 4, 27-41.

Stöber, J. (1995). Besorgnis: Ein Vergleich dreier Inventare zur Erfassung allgemeiner Sorgen [Worrying. A comparison of three questionnaires for assessment of everyday worries]. Zeitschrift für Differentielle und Diagnostische Psychologie, 16, 50-63.

Stöber, J. (1996). Besorgnis und Besorgtheit. Untersuchungen über Problemräume, deren Strukturierung und Elaboration [Worry and Worriers. Studies on problem spaces, their structuring and elaboration]. Frankfurt/M., Germany: Lang.

Stöber, J. (1997). Reliability and validity of two widely-used worry questionnaires: Self-reports and self-peer convergence. Manuscript submitted for publication.

Tallis, F., Davey, G. C. L., & Bond, A. (1994). The Worry Domains Questionnaire. In G. C. L. Davey & F. Tallis (Eds.), Worrying. Perspectives on theory, assessment, and treatment (pp. 285-297). New York: Wiley.

Tallis, F., Davey, G. C. L., & Capuzzo, N. (1994). The phenomenology of non-pathological worry: A preliminary investigation. In G. C. L. Davey & F. Tallis (Eds.), Worrying. Perspectives on theory, assessment, and treatment (pp. 61-89). New York: Wiley.

Tallis, F., Eysenck, M. W., & Mathews, A. (1991). Elevated evidence requirements and worry. Personality and Individual Differences, 12, 21-27.

Tallis, F., Eysenck, M. W., & Mathews, A. (1992). A questionnaire for the measurement of nonpathological worry. Personality and Individual Differences, 13, 161-168.

Wells, A. (1994). A multi-dimensional measure of worry: Development and preliminary validation of the Anxious Thoughts Inventory. Anxiety, Stress, and Coping, 6, 289-299.

Wisocki, P. A., Handen, B., & Morse, C. K. (1986). The Worry Scale as a measure of anxiety among homebound and community active elderly. Behavior Therapist, 9, 91-95.

Acknowledgments

The authors would like to thank Thomas D. Borkovec, Knut A. Hagtvet, Volker Hodapp, and an anonymous reviewer for their helpful comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to Jutta Joormann, Department of Research Methods, Social Psychology, and Developmental Psychology, Free University of Berlin, FB 12, Habelschwerdter Allee 45, D-14195 Berlin, Germany.

Table 1

The Item-Based Five-Factor Model. Standardized Factor Loadings and Uniqueness

	Domain[a]							
No.[b]	Rel	L of C	Aim Fut	Work Inc	Fin	SE[c]	Uniqueness	SE[d]
4.	.47 (.39)					.05 (.05)	.78 (.85)	.05 (.03)
16.	.59 (.47)					.05 (.05)	.65 (.78)	.05 (.05)
19.	.67 (.66)					.04 (.04)	.55 (.56)	.04 (04)
21.	.54 (.63)					.05 (.04)	.71 (.60)	.05 (.04)
23.	.81 (.73)					.04 (.04)	.34 (.47)	.04 (.04)
2.		.76 (.78)				.04 (.04)	.42 (.40)	.03 (.03)
10.		.82 (.85)				.04 (.04)	.34 (.28)	.03 (.03)
15.		.77 (.72)				.04 (.04)	.41 (.49)	.03 (.04)
18.		.66 (.63)				.04 (.04)	.57 (.61)	.04 (.04)
20.		.74 (.73)				.04 (.04)	.45 (.47)	.04 (.03)
3.			.47 (.56)			.05 (.04)	.78 (.69)	.05 (.05)
5.			.71 (.77)			.04 (.04)	.50 (.40)	.04 (.03)
8.			.43 (.58)			.05 (.05)	.81 (.72)	.06 (.05)
13.			.59 (.58)			.05 (.04)	.66 (.67)	.05 (.05)
22.			.74 (.71)			.04 (.04)	.45 (.50)	.04 (.04)
6.				.65 (.74)		.05 (.04)	.58 (.46)	.05 (.04)
14.				.63 (.76)		.05 (.04)	.60 (.43)	.05 (.04)
17.				.71 (.73)		.05 (.04)	.49 (.47)	.05 (.04)
24.				.41 (.34)		.05 (.05)	.83 (.89)	.06 (.06)
25.				.53 (.52)		.05 (.05)	.72 (.74)	.05 (.05)

(Table 1, continued)

1.	.75 (.68)	.04 (.04)	.44 (.54)	.04 (.04)
7.	.65 (.62)	.04 (.04)	.58 (.62)	.04 (.05)
9.	.84 (.82)	.04 (.04)	.30 (.33)	.03 (.04)
11.	.74 (.67)	.04 (.04)	.46 (.55)	.04 (.04)
12.	.57 (.58)	.05 (.04)	.68 (.67)	.05 (.05)

Note. First sample, N = 466. Values enclosed in parentheses represent values from second sample (N = 503).

[a]Rel = Relationships, L of C = Lack of Confidence, Aim Fut = Aimless Future, Work Inc = Work Incompetence, and Fin = Financial. [b]Item number (item wordings, see Appendix). [c]Standard errors of factor loadings. [d]Standard errors of uniquenesses.

Table 2

The Five-Factor Model. Intercorrelations of Factors, Descriptive Statistics, and Reliabilities

	Rel	L of C	Aim Fut	Work Inc	Fin
L of C	.79 (.78)				
Aim Fut	.79 (.76)	.75 (.71)			
Work Inc	.53 (.46)	.52 (.52)	.64 (.72)		
Fin	.23 (.42)	.24 (.22)	.43 (.44)	.38 (.33)	
M	4.99 (5.24)	5.93 (6.62)	5.59 (6.02)	6.32 (6.49)	4.69 (5.07)
SD	3.93 (3.68)	4.17 (4.19)	3.70 (3.96)	3.36 (3.69)	4.07 (3.80)
Cronbach's α	.75 (.71)	.86 (.85)	.70 (.76)	.72 (.75)	.83 (.80)

Note. First sample, N = 466. Values enclosed in parentheses represent values from the second sample (N = 503). For explanation of subscale abbreviations, see domains note of Table 1.

Table 3

Item-Based Models. Fit Indices for Comparison of Models of Different Factor Structure

Test	χ^2	df	RNI	PNFI	χ^2_{diff}	df_{diff}
Five-factor model	899.87** (944.71**)	265	.86 (.86)	.72 (.72)		
Three-factor model	1093.04** (1241.90**)	272	.82 (.80)	.71 (.69)		
Two-factor model	1296.64** (1477.46**)	274	.78 (.75)	.67 (.66)		
One-factor model	2002.84** (2061.02**)	275	.62 (.63)	.54 (.55)		
Null model	4858.51** (5158.47**)	300				
Five- vs. three-factor					193.17** (297.19**)	7
Three- vs. two-factor					203.60** (235.56**)	2
Three- vs. one-factor					706.20** (583.56**)	1

Note. First sample, N = 466. Values enclosed in parentheses represent values from second sample (N = 503). RNI = relative noncentrality index; PNFI = parsimonious normed fit index (cf. text for details).

**$p < .01$.

Table 4

Subscale Models. Fit Indices for Comparison of Models of Different Factor Structure

Test	χ^2	df	RNI	PNFI	χ^2_{diff}	df_{diff}
Five-factor model	89.61** (84.96**)	25	.97 (.98)	.53 (.54)		
Three-factor model	192.94** (230.37**)	32	.93 (.92)	.65 (.64)		
Two-factor model	319.25** (361.54**)	34	.87 (.86)	.65 (.64)		
One-factor model	634.07** (669.08**)	35	.72 (.73)	.56 (.56)		
Null model	2210.07** (2403.66**)	45				
Five- vs. three-factor					103.33** (145.41**)	7
Three- vs. two-factor					126.31** (131.17**)	2
Two- vs. one-factor					314.82** (307.54**)	1

Note. First sample, N = 466. Values enclosed in parentheses represent values from second sample (N = 503). RNI = relative noncentrality index; PNFI = parsimonious normed fit index.

**$p < .01$.

Table 5

Item-Based Models. Cross Validation of the Five-Factor Model

Test	χ^2	df	GoF	aGoF	χ^2_{diff}	df_{diff}
Model 1	944.71**	265	.87	.83		
Model 2	993.47**	285	.86	.84		
Model 3	1047.38**	300	.85	.84		
Model 1 vs. Model 2					48.76**	20
Model 2 vs. Model 3					53.91**	15

Note. Second sample, N = 503. Model 1 = Baseline model (no equality constraints); Model 2 = factor loadings fixed to be equal to the first sample; Model 3 = factor loadings and factor correlations fixed to be equal to the first sample. GoF = goodness of fit index; aGoF = adjusted goodness of fit index

**$p < .01$.

Table 6

Subscale Models. Cross Validation of the Five-Factor Model

Test	χ^2	df	GoF	aGoF	χ^2_{diff}	df_{diff}
Model 1	84.96**	25	.97	.93		
Model 2	90.82**	30	.97	.94		
Model 3	136.66	45	.95	.94		
Model 1 vs. Model 2					5.86[a]	5
Model 2 vs. Model 3					45.84**	15

Note. Second sample, N = 503. Model 1 = Baseline model (no equality constraints); Model 2 = factor loadings fixed to be equal to the first sample; Model 3 = factor loadings and factor correlations fixed to be equal to the first sample. GoF = goodness of fit index; aGoF = adjusted goodness of fit index

**$p < .01$; [a]$p > .25$.

Appendix

The Worry Domains Questionnaire

No.	Item wording: "I worry …"	Domain
4.	that my family will be angry with me or disapprove of something that I do	Rel
16.	that I find it difficult to maintain a stable relationship	Rel
19.	that I am unattractive	Rel
21.	that I will lose close friends	Rel
23.	that I am not loved	Rel
2.	that I cannot be assertive or express my opinions	L of C
10.	that I feel insecure	L of C
15.	that others will not approve of me	L of C
18.	that I lack confidence	L of C
20.	that I might make myself look stupid	L of C
3.	that my future job prospects are not good	Aim Fut
5.	that I'll never achieve my ambitions	Aim Fut
8.	that I have no concentration	Aim Fut
13.	that life may have no purpose	Aim Fut
22.	that I haven't achieved much	Aim Fut
6.	that I will not keep my workload up to date	Work Inc
14.	that I don't work hard enough	Work Inc
17.	that I leave work unfinished	Work Inc
24.	that I will be late for an appointment	Work Inc
25.	that I make mistakes at work	Work Inc
1.	that my money will run out	Fin
7.	that financial problems will restrict holidays and travel	Fin
9.	that I am not able to afford things	Fin
11.	that I can't afford to pay bills	Fin
12.	that my living conditions are inadequate	Fin

Note. The abbreviations for the domains are: Rel = Relationships, L of C = Lack of Confidence, Aim Fut = Aimless Future, Work Inc = Work Incompetence, and Fin = Financial. Item wordings and numberings are taken from Tallis, Davey, and Bond (1994, p. 288, Fig. 12.1).