Hungarian Validation of the Penn State Worry Questionnaire (PSWQ) – Comparing Latent Models with One or Two Method Factors Using both Paper-pencil and Online Versions of the PSWQ

ARTICLE *in* EUROPEAN JOURNAL OF PSYCHOLOGICAL ASSESSMENT · OCTOBER 2014

Impact Factor: 2.53 · DOI: 10.1027/1015-5759/a000221

Title: Hungarian Validation of the Penn State Worry Questionnaire (PSWQ) - Comparing Latent Models with One or Two Method Factors Using both Paper-pencil and Online Versions of the PSWQ

Short title: Hungarian Validation of the PSWQ

Summary

The Hungarian version of the Penn State Worry Questionnaire (PSWQ) was validated in two studies, using five different samples. Study 1 tested the factor structure and internal consistency of the PSWQ in two undergraduate student samples, comparing the psychometric properties of the paper-pencil and the online versions of the scale. Study 2 assessed construct validity in two undergraduate student samples and in a sample of patients diagnosed with Generalized Anxiety Disorder (GAD) and matched control participants. Our results suggest that the Hungarian PSWQ demonstrates good psychometric properties. We found no difference between the online and the paper-pencil versions of the scale. A factor structure with one general worry factor and two method factors representing wording effects showed the best fit to the data.

Keywords: worry, Penn State Worry Questionnaire, factor structure, method factors, online administration

Hungarian Validation of the Penn State Worry Questionnaire (PSWQ) - Comparing Latent Models with One or Two Method factors Using both Paper-pencil and Online Versions of the PSWQ

Worry refers to "a chain of thoughts and images, negatively affect-laden and relatively uncontrollable" (Borkovec, Robinson, Pruzinsky, & DePree, 1983, p. 10). It is prevalent in many mood and anxiety disorders (Purdon & Harrington, 2006). Moreover, chronic, excessive and uncontrollable worrying is a core diagnostic feature of Generalized Anxiety Disorder (GAD, American Psychiatric Association, 2001).

One popular instrument measuring pathological worry is the Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990). It consists of 16 items, focusing on the excessiveness and uncontrollability of the worry process. Subjects are required to respond on 5-point Likert scales ranging from 1 (*not at all typical*) to 5 (*very typical*). Eleven items are positively worded with higher ratings reflecting more pronounced worrying (e.g. item 2: "My worries overwhelm me"). The other five items are negatively worded with higher ratings reflecting the absence of worry (e.g. item 10: "I never worry about anything"). In the following, we review psychometric work related to the scale, focusing on three issues: basic psychometric properties, factor structure and the equivalence of paper-pencil and online versions of the scale.

Startup and Erickson (2006) summarizes the wealth of research demonstrating good psychometric properties of the PSWQ: it possesses adequate internal consistency (Cronbach's alpha = .80-.95), test-retest reliability (r = .74-.92) and construct validity, which is proved by strong correlations with measures of trait anxiety (r = .64-.79) and by somewhat weaker correlations with measures of depression (r = .36-.62).

This is in line with the original definition linking worry to the fear process (Borkovec et al., 1983), and also supports the view that depression and anxiety are related phenomena, sharing some but not all underlying processes (e.g. Clark & Watson, 1991). The PSWQ assesses pathological worrying characteristic of GAD, thus studies showing that GAD-patients score higher on the scale than matched controls suggest that criterion validity of the scale is appropriate (see Startup & Erickson, 2006).

Regarding the factor structure of the PSWQ, early studies using exploratory factor analysis (EFA) indicated two factors with positively and negatively worded items loading on distinct factors. Because of this, some investigators (e.g. Beck, Stanley, & Zebb, 1995; Stöber, 1995) retained a two-factor solution with two distinct but correlated factors, representing the positively and the negatively worded items, respectively. In contrast, due to the unitary theoretical construct of worry and the high internal consistency of the scale, others (e.g. Meyer et al., 1990; van Rijsoort, Emmelkamp, & Vervaeke, 1999) favoured a one-factor solution, with all 16 items loading on one general factor.

Brown (2003) and Hazlett-Stevens, Ullman, and Craske (2004) suggested a solution for this controversy: Using confirmatory factor analysis (CFA), they demonstrated that the distinct factors appearing in EFA might reflect method effects caused by different responding to positively and negatively worded items. Thus all items of the PSWQ measure the unitary theoretical construct of worry, but additionally, some of the items share additional common variance (related to the wording of these items).

For such cases, Reise, Moore, and Haviland (2010) suggested the use of bifactor models, in which items load on one general factor and also on an orthogonal method factor. Consistently, recent adaptations of the PSWQ found best fit for bifactor models, which consist of either one general trait factor combined with one method factor for the negatively worded items (e.g. Lim, Kim, Lee, & Kwom, 2008; van der Heiden, Muris, Bos, & van der

Molen, 2010), or one general trait factor combined with two distinct method factors for the positively and the negatively worded items, respectively (e.g. Gana, Martin, Canouet, Trouillet, & Meloni, 2002; Pallesen, Nordhus, Carlstedt, Thayer, & Johnsen, 2006).

Finally, two studies examined the psychometric properties of the PSWQ administered online. Zlomke (2009) found low, whereas Verkuil and Brosschot (2012) found appropriate internal consistency. Both studies reported a high total score of the scale, as compared to the paper-pencil version. This might be due to higher self-disclosure associated with online administration (Buchanan, 2003), or in the case of Verkuil and Brosschot’s study, the authors suggest that participation could have been more appealing for high worriers, because the study was explicitly advertised as focusing on worry.

The aim of the present study was threefold. First, we aimed to construct the Hungarian adaptation of the PSWQ and to examine its basic psychometric properties: construct validity, internal and test-retest reliability. Second, we wanted to investigate the factor structure of the Hungarian PSWQ to test which of the previously described latent models fits the data best on a Hungarian sample. Third, we aimed to find evidence for the equivalence of the online and the paper-pencil version of the Hungarian translation of the PSWQ regarding internal reliability, factor structure and total score.

The latter two aims require complex CFA analyses and thus large sample sizes, but only the items of the PSWQ have to be administered. In contrast, establishing construct validity requires the assessment of multiple scales but is less reliant on large sample sizes. Thus, to optimize the utilization of research resources, we conducted two studies: In Study 1, factor structure and reliability were assessed for both the paper-pencil and online administration using two large undergraduate samples. Thereafter, in Study 2, we examined construct validity using a larger set of questionnaires in two undergraduate samples of moderate size and in a small sample consisting of GAD-patients and matched control participants. In accordance with previous findings, we expected strong correlation with trait anxiety and a somewhat weaker correlation with depression. As a further measure of construct validity, punishment and reward sensitivity was assessed. Torrubia, Ávila, Molto, and Caseras (2001) found that the former, but not the latter is correlated with trait anxiety, thus we expected the same pattern for the PSWQ.

Translation Process

The original items were translated by the first author, and were checked by a research assistant for linguistic correctness. After that, two bilingual translators translated the items back to English which were than compared to the original items. Before finalizing the Hungarian items, minor changes were made based on the comparison of the original and the back translated items. (The translated items are available from the first author upon request.)

Study 1- Factor Structure, Reliability and Convergent Validity

Materials and Methods

Samples and data collection. Data for Sample 1 (N = 702, 316 women, M_{age} = 21.52, SD_{age} = 1.98, age range = 19-34) were collected in a one-year period, during which participants appearing in our lab for participating in other experiments filled out the PSWQ. These experiments targeted memory or executive functions, thus it is unlikely that they influenced PSWQ scores. A small subset of the participants filled out the PSWQ again three weeks later for establishing test-retest reliability (N = 42, 16 women, M_{age} = 21.19, SD_{age} = 1.71, age range = 19-27).

In Sample 2 (N = 637, 477 women, M_{age} = 20.10, SD_{age} = 1.59, age range = 18-29), the PSWQ was administered online, as part of a large questionnaire battery aimed to screen participants for an ongoing experiment in our laboratory.

Participants were in both cases undergraduate students participating for partial credit in psychology courses. They signed the consent form including information about the anonymous nature of the study. Full anonymity, however, could not be fulfilled in Sample 2, because participants entered their email address.

Data analysis. We used CFA to compare the four models described earlier in the literature: (1) a single-factor model with all 16 items loading on one factor (M1, e.g. Meyer et al., 1990); (2) a two-factor model with two latent factors representing the positively and the negatively worded items, respectively (M2, e.g. Stöber, 1995); (3) a bifactor model consisting of a general trait factor representing all 16 items and one method factor representing the negatively worded items (M3, e.g. Hazlett-Stevens et al., 2004); (4) a bifactor model with one general trait factor and two method factors representing the positively and the negatively worded items, respectively (M4, e.g. Gana et al., 2002). The latent factors were correlated only in M2, and uniqueness correlations were set to zero in all models.

We also investigated measurement invariance (MI) across administration modes, that is, we aimed to show that items measure the same theoretical construct in the same way in the online and in the paper-pencil versions of the scale. First, we examined whether the same CFA model shows best fit in both samples. Thereafter, the equality of different model parameters was tested using multi-group CFA (MGCFA), with the two samples pooled together, as two groups of the same sample. In MGCFA, MI is tested by investigating changes in model fit after specific model parameters have been fixed to be invariant between groups. We aimed to test the invariance of factor loadings and intercepts, because this form of invariance, called scalar invariance, is required to meaningfully compare latent or observed means across groups (Steinmetz, 2013). It can be established by first testing the fit of a configural model, in which no model parameter is invariant. This model is then compared with a scalar model, in which factor loadings and intercepts are fixed to be invariant between the groups. Scalar invariance is established if the fit of these two models do not significantly differ.

CFA and MGCFA were conducted using the statistical modelling software Mplus (Version 6.11; Muthén & Muthén, 1998-2010). We used a robust maximum-likelihood estimator (MLR; Muthén & Muthén, 1998-2010) and a mean and variance adjusted weighted-least square estimator (WLSMV; Muthén & Muthén, 1998-2010). Both can be used for five-category ordinal indicators (Rhemtulla, Brosseau-Liard, & Savalei, 2012). WLSMV is specifically designed for ordinal indicators, whereas MLR allows more flexibility in model comparisons. For CFA and MGCFA, negatively worded items were reverse scored.

To determine model fit, we used the χ^2 statistics. Significant χ^2 indicates that the model does not fit the data well. With large sample sizes and complex models, however, this test is too conservative (i.e. too fast rejecting the model), thus we also investigated the comparative fix index (CFI) and the root mean square error of approximation (RMSEA). Based on Yu (2000), the following criteria were adopted for well-fitting model: CFI $\geq$ 0.96 and RMSEA $\leq$ 0.05. We also computed the confidence interval of RMSEA. A model fits well when the lower bound of the 90% confidence interval for RMSEA is below 0.05 and the upper bound is below 0.08.

Relative fit of different CFA models was compared by investigating the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These are information theory-based fit indices with lower values indicating better fit. The scalar and configural models in MGCFA were compared by investigating the difference in the χ^2 values ($\Delta\chi^2$). This χ^2 difference-testing, however, is also sensitive to sample size and model complexity. Thus for testing MI, Cheung and Remswood (2002) suggests to investigate decrement in other fit indices: ΔCFI > 0.01 and ΔRMSEA > 0.015 indicate significant deterioration in model fit. These cut-offs and the AIC/BIC values are only available and reported for the MLR estimator.

Besides, observed means and internal reliability were compared across the two administration modes and the correlation between the PSWQ scores at baseline and follow-up was assessed in the test-retest subsample of Sample 1.

Results

Descriptive statistics and reliability analysis. The mean PSWQ score in the two samples (43.36 (SD = 12.73) in Sample 1 and 43.18 (SD = 13.01) in Sample 2) did not differ significantly, $t(1337) = 0.26$, $p = .79$. Cronbach's alpha values indicated excellent internal consistency in both samples (Sample 1: $\alpha = .93$; Sample 2: $\alpha = .94$). Item-total correlations were between .60 and .80 for most items. A relatively lower, but still acceptable item-total correlation was observed for item 1 (Sample 1: $r_{it} = .39$; Sample 2: $r_{it} = .50$), item 11 (Sample 1: $r_{it} = .51$; Sample 2: $r_{it} = .52$) and item 16 (Sample 1: $r_{it} = .58$; Sample 2: $r_{it} = .59$). Test-retest reliability tested in the subsample of Sample 1 was adequate ($r_s = .87$, $p < .001$).

Factor structure and measurement invariance. For CFA, goodness of fit indices are presented in Table 1 for both estimators. χ^2 values were significant for all models, possibly due to large sample sizes. Fit indices indicate that M4 fits the data best in both samples: it has the lowest BIC, AIC, RMSEA and the highest CFI values. Moreover, M4 is the only model satisfying the cut-offs for a well-fitting model in both samples: CFI is above 0.96, RMSEA value approaches 0.05 and the lower bound of the 90% confidence interval for RMSEA is below 0.05 and the upper bound is below 0.08. Because of its superior fit, M4 was retained as the factor structure of the Hungarian PSWQ. Factor loadings for M4 were similar for both estimators, thus only the loadings estimated with WLSMV are presented in Table 2. Factor loadings of the method factor representing positively worded items are weak and inconsistent, whereas loadings on the other two factors are satisfactory. This suggests weaker method-effects for positively than for negatively worded items.

Because M4 showed the best fit in both samples, it was chosen to test measurement invariance of the paper-pencil and the online versions using MGCFA. Estimated with MLR, the configural model showed appropriate fit ($\chi^2(176) = 443$, $p < .001$, CFI = 0.974, RMSEA = 0.048 [0.042, 0.053]) and the same was true for the more constrained scalar model ($\chi^2(218) = 548$, $p < .001$, CFI = 0.968, RMSEA = 0.048 [0.043, 0.053]). Whereas $\Delta\chi^2$ indicate that the scalar model shows a worse fit than the configural model ($\Delta\chi^2(42) = 105$, $p < .001$), ΔCFI and ΔRMSEA suggest no difference (ΔCFI = 0.006; ΔRMSEA < 0.001). Because $\Delta\chi^2$-tests are sensitive to sample size, we based our decision on ΔCFI and ΔRMSEA. Both values were below the cut-off suggested by Cheung and Remswood (2002), thus we concluded that scalar invariance was established.

Study 2- Construct Validity

Materials and Methods

Samples and Data Collection. Participants of the two undergraduate samples were recruited and informed as described for Sample 1 of Study 1. Subjects participating in experiments in our memory lab filled in either a shorter or a longer questionnaire packet on paper, depending on the duration of the experiment they participated in. Thus data were collected for two samples.

In Sample 1 (N = 126, 56 women, $M_{age} = 24.01$, $SD_{age} = 5.73$, age range = 18-42), besides the PSWQ, we assessed the trait subscale of the Spielberger State-Trait Anxiety Inventory (STAI-T; Spielberger, Gorsuch, & Lushene, 1970; Hungarian version: Sipos, Sipos, & Spielberger, 1998) which consists of 20 items, answered on 4-point Likert scales, assessing the proneness to common signs and symptoms of anxiety. Additionally, we administered the Sensitivity to Punishment and Sensitivity to Reward Scale (SPSRQ; Torrubia et al., 2001; Hungarian version: Kállai, Rózsa, Kerekes, Hargitai, & Osváth, 2009) which consists of 48 items answered on a dichotomous scale (*typical* – *not typical*) and has two subscales: Sensitivity to Punishment and Sensitivity to Reward.

In Sample 2 (N = 151, 88 women, M_{age} = 22.97, SD_{age} = 3.08, age range = 18-33), we administered PSWQ, STAI-T and a short 9-item version of the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) which has the same psychometric properties as the original version of the scale (Rózsa, Szádóczky, & Füredi, 2001). Items of the BDI focus on signs of depression and are scored on 4-point Likert scales.

Additionally, in Sample 3, we investigated the criterion validity of the Hungarian PSWQ by administering it to a group of outpatients diagnosed with GAD (N = 13, 10 women, M_{age} = 45.54, SD_{age} = 14.19, age range = 29-72, $M_{education}$ = 14.31; $SD_{education}$ = 2.59) and compared their PSWQ score to a group of control participants matched for age, gender and education (N = 13, 10 women, M_{age} = 42.15, SD_{age} = 10.03, age range = 30-62, $M_{\text{education}}$ = 12.77; $SD_{\text{education}}$ = 6.14). Both patients and controls gave informed consent.

Data analysis. In Sample 1 and 2, the total scores of the scales were non-normally distributed, thus the association between them was examined by computing the Spearman correlation coefficient. In Sample 3, the PSWQ score was normally distributed in both groups, thus the group means were compared by t-test.

Results

Mean PSWQ score was 41.56 (SD = 11.13) in Sample 1 and 46.19 (SD = 12.97) in Sample 2. Correlation analysis indicated that worry was strongly associated with trait anxiety in both samples (Sample 1: $r_s = .75$, $p < .001$, Sample 2: $r_s = .72$, $p < .001$). The correlation coefficient yielded a weaker link both between worry and punishment sensitivity in Sample 1 ($r_s = .52$, $p < .001$) and between worry and depression in Sample 2 ($r_s = .38$, $p < .001$). Worry was not associated to reward sensitivity in Sample 1 ($r_s = .13$, $p = .14$).

In Sample 3, GAD-patients scored significantly higher on the PSWQ than matched controls (GAD-patients: $M = 59.00$, $SD = 11.58$; controls: $M = 43.38$, $SD = 10.02$; $t(24) = 3.68$, $p < .01$).

Discussion

The first aim of this study was to investigate basic psychometric properties of the Hungarian PSWQ. Results of Study 1 confirmed that the Hungarian PSWQ has the same excellent internal consistency as the original English PSWQ. In a small subsample, we also demonstrated good test-retest reliability of the scale over a period of three weeks. In Study 2, construct validity was demonstrated by showing that the Hungarian PSWQ relates to other constructs in a theoretically meaningful way: it was strongly related to trait anxiety, moderately to punishment sensitivity and depression, whereas the association between worry and reward sensitivity was not significant. Finally, GAD-patients had significantly higher PSWQ scores than matched controls, demonstrating criterion validity.

Our second aim was to examine the factor structure of the Hungarian PSWQ. Using CFA, we replicated previous research (e.g. Lim et al., 2008; Pallesen et al., 2006) by showing that bifactor models with trait and method factors outperformed models with one general or two distinct, but correlated trait factors. We also showed that a bifactor model with only one method factor for the negatively worded items is outperformed by a bifactor model with distinct method factors for positively and negatively worded items. The factor loadings of the positively worded method factor, however, were inconsistent and weak. Such a pattern was also observed in other studies (Gana et al., 2002; Pallesen et al., 2006), thus it seems to be inherent to the scale, and is not related to the Hungarian version. Our results indicate, that these inconsistent loadings notwithstanding, taking method effects related to positively worded items into account significantly enhances model fit, at least in the case of the

Hungarian PSWQ. Therefore, it may be advised to retain a model with one trait and two method factors.

Finally, our third aim was to compare the online and the paper-pencil versions of the scale. We found no differences regarding internal reliability and factor structure. We established scalar invariance suggesting that total scores gathered via the two administration modes can be meaningfully compared. Finally, in contrast with previous investigators (Verkuil & Brosschot, 2012; Zlomke, 2009), we did not observe elevated total score of the scale administered online. This might have been caused by the fact that, in contrast to the study of Verkuil and Brosschot (2012), in our study high worriers were not likely to be overrepresented; the aim of the study was not the investigation of worry specifically. Alternatively, asking for participants' email address during the online administration impeded self disclosure to the same extent as did personal contact with the experimenter in the laboratory setting, resulting in similar total scores.

The present study has important limitations. First, we used undergraduate samples, thus one should be cautious when generalizing our results to other age groups. Second, we used relatively small samples for testing test-retest reliability in Study 1 and criterion validity in Study 2. Third, we did not assess convergent validity of the scale by using another questionnaire focusing on some form of repetitive thinking (Watkins, 2008). And finally, we did not investigate the construct validity of the online administered version of the PSWQ.

Notwithstanding these limitations, our results suggest that both the paper-pencil and the online versions of the Hungarian PSWQ are reliable and valid measures of pathological worry. The latent structure of the Hungarian PSWQ consists of one general trait factor measuring pathological worrying and two method factors related to the wording of the items. Our study also demonstrated that this latent structure shows superior fit to a solution with only one method factor.

Acknowledgement

This work was supported by the Hungarian Scientific Research Found (OTKA K84019) and by the Early Career Stimulus Award awarded to the first author by the European Society for Cognitive Psychology.

References

American Psychiatric Association (2001). *Diagnostic and statistical manual of mental disorders—Text revision* (4th ed.). Washington, DC: APA.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

Beck, J.G., Stanley, M., & Zebb, B.J. (1995). Psychometric properties of the Penn State Worry Questionnaire in older adults. *Journal of Clinical Geropsychology, 1*(1), 33-42.

Brown, T.A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy, 41*(12), 1411–1426

Borkovec, T.D., Robinson, E., Pruzinsky, T., & DePree, J.A. (1983). Preliminary exploration of worry: Some characteristics and processes. *Behaviour Research and Therapy, 21*(1), 9–16.

Buchanan, T. (2003). Internet-based questionnaire assessment: appropriate use in clinical contexts. *Cognitive Behaviour Therapy, 32*(3), 100–109.

Cheung, G.W., & Remswood, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233-255.

Clark, L.A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316-336.

Gana, K., Martin, B., Canouet, M-D., Trouillet, R., & Meloni, F. (2002). Factorial structure of a French version of the Penn State Worry Questionnaire. *European Journal of Psychological Assessment, 18*(2), 158–164.

Hazlett-Stevens, H., Ullman, H.B., & Craske, M.G. (2004). Factor structure of the Penn State Worry Questionnaire: Examination of a method factor. *Assessment, 11*(4), 361-370.

Kállai, J., Rózsa, S., Kerekes, Zs., Hargitai, R., & Osváth, A. (2009). A temperamentum alapdimenziójának egyik mérőeszközével, a büntetés- és jutalomérzékenység kérdőívvel (SPSRQ) szerzett hazai tapasztalatok. [A temperament dependent dimension: psychometric properties of the Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) on Hungarian sample]. *Mentálhigiéné és Pszichoszomatika, 10*(2), 153-172.

Lim, Y-J., Kim, Y-H., Lee, E-H., & Kwon, S-M. (2008). The Penn State Worry Questionnaire: psychometric properties of the Korean version. *Depression and Anxiety 25*(10), E97-E103.

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*(6), 487–495.

Muthén, L.K., & Muthén, B.O. (1998-2010). *Mplus user's guide* (6th edition). Los Angeles, CA: Muthén & Muthén.

Pallesen, S., Nordhus, I.H., Carlstedt, B., Thayer, J.F., & Johnsen, T.B. (2006). A Norwegian adaptation of the Penn State Worry Questionnaire: Factor structure, reliability, validity and norms. *Scandinavian Journal of Psychology, 47(4),* 281–291.

Purdon, C., & Harrington, J. (2006). Worry in psychopathology. In G.C.L. Davey & A. Wells (Eds.), *Worry and its psychological disorders: Theory, assessment and treatment* (pp. 41-51). West Sussex, England: Wiley and Sons.

Reise, S.P., Moore, T.M., & Haviland, M.G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559.

Rhemtulla M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373.

Rózsa S., Szádóczky E., & Füredi J. (2001). A Beck Depresszió Kérdőív rövidített változatának jellemzői a hazai mintán [Psychometric properties of the shortened version of the Beck Depression Inventory on a national sample]. *Psychiatria Hungarica, 16*(4), 379-397.

Sipos K., Sipos M., & Spielberger C.D. (1998). A State-Trait Anxiety Inventory (STAI) magyar változata [The Hungarian version of the State-Trait Inventory (STAI)]. In F. Mérei & F. Szakács F. (Eds.). *Pszichodiagnosztikai vademecum* [Handbook for Psychodiagnostics] (Vol 1., pp 123-135). Budapest, Hungary: Nemzeti Tankönyvkiadó.

Spielberger, C.D., Gorsuch, R.L., & Lushene, R.E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists' Press.

Startup, H.M., & Erickson, T.M. (2006). The Penn State Worry Questionnaire. In: G.C.L. Davey & A. Wells: *Worry and its psychological disorders. Research, theory and treatment.* (pp. 101-121) West Sussex, England: Wiley and Sons.

Steinmetz, H. (2013). Analyzing observed composite differences across groups. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9*(1), 1-12.

Stöber, J. (1995). Besorgnis: Ein Vergleich dreier Inventare zur Erfassung allgemeiner Sorgen. [Worrying: A comparison of three questionnaires concerning everyday worries]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 16*(1), 50-63.

Torrubia, R., Avila, C., Moltó, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences, 31*(6), 837-862.

van der Heiden, C., Muris, P., Bos, A.E.R., & van der Molen, H.T. (2010). Factor structure of the Dutch version of the Penn State Worry Questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 41*(3), 304-309.

van Rijsoort, S., Emmelkamp, P., & Vervaeke, G. (1999). The Penn State Worry Questionnaire and the Worry Domains Questionnaire: structure, reliability and validity. *Clinical Psychology and Psychotherapy, 6*(4), 297-307.

Verkuil, B., & Brosschot, J.F. (2012). The online version of the Dutch Penn State Worry Questionnaire: Factor structure, predictive validity and reliability. *Journal of Anxiety Disorders, 26*(8), 844–848.

Watkins, E.R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin*, *134*(2), 163-206.

Yu, C.Y. (2002). *Evaluation of cut-off criteria of model fit indices for latent variable models with categorical and continuous outcomes* (Doctoral dissertation, University of California, Los Angeles). Retrieved from http://www.statmodel.com/download/Yudissertation.pdf

Zlomke, K.R. (2009). Psychometric properties of internet administered versions of Penn State Worry Questionnaire (PSWQ) and Depression, Anxiety, and Stress Scale (DASS). *Computers in Human Behavior, 25*(4), 841–843.

Table 1

Fit indices of the latent models in Study 1 for the paper-pencil (Sample 1) and the online version (Sample 2) of the Hungarian PSWQ.

	Model	χ^2(df)[a]	*CFI*	RMSEA - 90% CI	AIC	BIC
WLSMW, Sample 1	M1	762(104)*	0.968	0.095 [0.089, 0.101]		
	M2[a]	360(103)*	0.987	0.060 [0.053, 0.066]		
	M3	335(99)*	0.988	0.058 [0.051, 0.065]		
	M4	247(88)*	0.992	0.051 [0.043, 0.058]		
WLSMV, Sample 2	M1	718(104)*	0.966	0.096 [0.090, 0.103]		
	M2[a]	414(103)*	0.983	0.069 [0.062, 0.076]		
	M3	376(99)*	0.985	0.066 [0.059, 0.073]		
	M4	264(88)*	0.990	0.056 [0.048, 0.064]		
MLR, Sample 1	M1	613(104)*	0.903	0.083 [0.077, 0.090]	28,953	29,172
	M2[a]	367(103)*	0.950	0.060 [0.054, 0.067]	28,666	28,889
	M3	345(99)*	0.953	0.060 [0.053, 0.066]	28,650	28,891
	M4	234(88)*	0.972	0.049 [0.041, 0.056]	28,525	28,817
MLR, Sample 2	M1	616(104)*	0.902	0.088 [0.081, 0.095]	25,779	25,993
	M2[a]	381(103)*	0.947	0.065 [0.058, 0.072]	25,517	25,736
	M3	348(99)*	0.952	0.063 [0.056, 0.070]	25,490	25,726
	M4	209(88)*	0.977	0.047 [0.038, 0.055]	25,353	25,638

Note. χ^2(df): χ^2-value with degrees of freedom; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; CFI: Comparative Fit Index; RMSEA - 90% CI: Root Mean Square Error of Approximation with 90% confidence intervals; MLR: robust maximum-likelihood estimator; WLSMV: mean and variance adjusted weighted-least square estimator; M1: model with one general trait factor; M2: model with two correlated latent factors for the positively and the negatively worded items; M3: model with one global trait factor and one method factor for the negatively worded items; M4: model with one global trait factor and two method factors for the positively and the negatively worded items, respectively.

* $p < .001$

[a] Correlation of the two latent factors stronger then -.75, $p < .001$

Table 2

Standardized factor loadings of the PSWQ items in the two samples of Study 1 for the latent model with one trait and two method factors (M4).

	Trait Factor		Positively worded method factor		Negatively worded method factor	
	S1	S2	S1	S2	S1	S2
item 1	-.39	-.52			.47	.44
item 2	.73	.71	-.09†	.15		
item 3	-.64	-.76			.50	.35
item 4	.82	.82	-.17*	.12*		
item 5	.78	.79	-.20	-.04†		
item 6	.73	.81	-.07†	-.02†		
item 7	.88	.85	.18	.26		
item 8	-.64	-.72			.32	.21
item 9	.68	.62	.37	.21		
item 10	-.70	-.76			.49	.45
item 11	-.51	-.53			.33	.38
item 12	.83	.83	.01†	.20		
item 13	.85	.77	-.02†	.28		
item 14	.77	.73	.04†	.32		
item 15	.94	.82	.15*	.56		
item 16	.61	.64	.25	.11*		

Note. S1: Sample 1, paper-pencil administration; S2: Sample 2, online administration. Factor loadings with no value are restricted to be equal to zero. For loadings not marked with symbol: $p < .001$

* $p < .05$

† Not significant