Short title: Generalised Pliance Questionnaire in Spain

MANUSCRIPT ACCEPTED FOR PUBLICATION IN EUROPEAN JOURNAL OF PYSCHOLOGICAL ASSESSMENT

Psychometric properties of the Generalised Pliance Questionnaire in Spain

Francisco J. Ruiz[1] (franciscoj.ruizj@konradlorenz.edu.co)

Paula Odriozola-González[2] (paula.odriozola.gonzalez@uva.es)

Juan C. Suárez-Falcón[3] (jcsuarez@psi.uned.es)

[1]Fundación Universitaria Konrad Lorenz, [2]Universidad de Valladolid, [3]Universidad Nacional de Educación a Distancia (UNED)

Correspondence address: Juan C. Suárez-Falcón, Facultad de Psicología, UNED, Ciudad Universitaria, C/ Juan del Rosal, 10, 28040 Madrid, jcsuarez@psi.uned.es

Authors' contribution: FJR, POG, and JCS designed the study; POG collected the data; JCS conducted data analyses; FJR and POG wrote the manuscript; JCS reviewed the manuscript.

Abstract

Rule-governed behaviour is a central concept to explain complex human behaviour from a functional-analytic standpoint. Recently, a self-report measure of a functional type of rule-following called generalised pliance has been developed – the Generalised Pliance Questionnaire (GPQ). The GPQ has two versions with 18 and 9 items, respectively. This study aimed to analyse the factor structure and psychometric properties of both GPQ versions in a Spanish online sample (N = 846). Confirmatory factor analyses showed that the one-factor model obtained an acceptable fit in the GPQ-18, but an inadequate fit in the GPQ-9. The GPQ-18 showed measurement invariance across gender, excellent internal consistency, and convergent construct validity. According to the results, the GPQ-18 showed good psychometric properties in Spanish samples.

Key words: Rule-governed behaviour; Generalised pliance; Relational frame theory; Acceptance and commitment therapy.

Psychometric properties of the Generalised Pliance Questionnaire in Spain

Rule-governed behaviour (RGB) is behaviour controlled by antecedent verbal stimuli and has been a key functional-analytic concept to explain complex human behaviour (Skinner, 1966). Several approaches for the explanation of RGB have been developed during the last decades. Among them, relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001) is probably the most complete approach because it provides explanations of the core characteristics and behavioural processes involved in rule-governed behaviour, including the conceptualisation of verbal stimuli, the generation, meaning and understanding of rules, and rule-following (Hayes & Hayes, 1989). Regarding rule-following, RFT has adopted the functional classification provided by Zettle and Hayes (1982): pliance, tracking, and augmenting. For the sake of brevity, in this paper, we will focus only on pliance and tracking.

The main difference between pliance and tracking is the *apparent* source of reinforcement for rule-following: social or arbitrary in the case of pliance and non-arbitrary in the case of tracking (i.e., following the rule directly causes the consequence). Importantly, the word "apparent" highlights that rules are antecedent events. Consequently, the classification of a rule as an instance of pliance (i.e., ply) or tracking (i.e., track) is only related to the functions actualised on the listener when performing the rule and not the actual consequences contacted when performing it. In other words, the present value of the rules is determined by the listener's history (Hayes & Hayes, 1989).

Pliance is rule-following due to a history of multiple exemplars in which the speaker provided the listener with reinforcement contingent on the correspondence between the rule content and the listener's behaviour (Luciano, Valdivia-Salas, & Ruiz, 2012; Zettle & Hayes, 1982). An individual follows a ply when he or she behaves

influenced by the rule to obtain socially-mediated consequences. For instance, a boy might follow his mother's rule, "Please, tidy your room" because he wants to avoid his mother's anger for not obeying. In this case, the rule worked as a ply because the boy follows it to prevent a socially-mediated consequence of punishment.

Pliance is the first functional class of rule-following developed because of its relational simplicity (Hayes, Zettle, & Rosenfarb, 1989; Törneke, Luciano, & Valdivia-Salas, 2008). Tracking usually develops after some practice with pliance if the child is guided to verbally contact the natural consequences of his or her behaviour (Luciano, Valdivia-Salas, Cabello, & Hernández, 2009). For instance, following the example above, the mother could guide the boy to realise that, after tidying the room, he will have more space to play. However, if these interactions are lacking, pliance will become the predominant type of rule-following for the child (Luciano et al., 2012; Törneke et al., 2008). This is the breeding ground for social approval to become the primary source of reinforcement for the individual, which is a pattern of RGB called generalised pliance (Ruiz, Suárez-Falcón, Barbero-Rubio, & Flórez, 2019). Generalised pliance has been theoretically associated with psychopathology because it precludes the verbal construction of values and contact with the natural consequences of one's own behaviour (Hayes, Strosahl, & Wilson, 1999). The latter makes individuals displaying generalised pliance particularly insensitive to contingencies (Hayes et al., 1989; Törneke et al., 2008).

The empirical analysis of generalised pliance has been facilitated by the development of the Generalised Pliance Questionnaire (GPQ; Ruiz et al., 2019). The GPQ is an 18-item measure that was developed in Colombia to be administered to adult, clinical and nonclinical samples. A short, 9-item version (i.e., GPQ-9) was developed in the same series of studies to be applied in survey research where the number of items of

the GPQ might be an issue. The GPQ-9 was designed by removing nine items based on the following criteria: (a) factor loadings and corrected item-total correlations, (b) avoiding overlapping content across items, and (c) avoiding items with content specifying negative private experiences to prevent potential contamination with items measuring negative affect.

Both versions of the GPQ showed excellent internal consistency and a one-factor structure. Importantly, the GPQ showed positive correlations with experiential avoidance, cognitive fusion, obstruction in values, dysfunctional schemas, and emotional symptoms, and negative correlations with progress in values, life satisfaction, and mindfulness skills. The GPQ-9 also showed treatment sensitivity in the randomised clinical trial conducted by Ruiz et al. (in press) that analysed the efficacy of a 2-session ACT protocol for the treatment of depression and generalised anxiety disorder.

The study by Waldeck, Pancani, and Tyndall (2019) found that the English translation of the GPQ-9 showed good psychometric properties and a one-factor structure. Additionally, the GPQ-9 showed strong correlations with measures of fear of negative evaluation and need to belong. O'Connor, Byrne, Ruiz, and McHugh (2019) showed in an Irish sample that the GPQ-18 has criterion validity according to the strong correlations found with tests of cognitive flexibility and insensitivity to contingencies such as the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948). Lastly, Harte et al. (in press) used the GPQ-18 in basic research on RGB in which they explored the potential impact of relational coherence on persistent rule-following. In the two studies conducted, the authors observed a positive and statistically significant correlation between the GPQ-18 and rule persistence (i.e., insensitivity to contingencies).

Salazar, Ruiz, Flórez, and Suárez-Falcón (2018) developed a version of the GPQ appropriate for children and adolescents (i.e., GPQ-C). This version also had good

psychometric properties and a one-factor structure. Likewise, Salazar, Ruiz, García-Martín, and Bedoya-Valderrama (submitted) showed that the GPQ-C can discriminate between children with learning difficulties and control participants and that GPQ-C scores showed a strong correlation with the performance on the WCST. Lastly, Stapleton and McHugh (2020) have also found that scores on the GPQ-C correlated positively with measures of empathic concern and capacity for interpersonal pleasure.

The review presented in the last paragraphs shows that research on generalised pliance is growing rapidly. However, a limitation of the research conducted to date is that the GPQ has only been analysed in Colombia, England, and Ireland. Accordingly, this study aimed to analyse the factor structure, measurement invariance across gender, internal consistency, and convergent construct validity of the GPQ in an online sample of 846 participants from Spain.

Method

Participants

The sample consisted of 846 participants (78.7% females) with age ranging between 18 and 72 years (M = 35.38, SD = 11.12). They responded to an anonymous internet survey distributed through social media. All of them were Spaniards. The relative educational level of the participants was: 0.1% no studies, 3.3% primary studies (i.e., compulsory education), 29.7% mid-level study graduates (i.e., high school or vocational training), and 65.6% were undergraduates or college graduates (1.3% did not indicate the educational level).

Instruments

Generalised Pliance Questionnaire (GPQ; Ruiz et al., 2019). The GPQ is a measure of generalised pliance that consists of 18 items that are responded to on a 7-point Likert-type scale (1 = *never true*; 7 = *always true*). A shorter, 9-item version was

also developed in the same validation study. Both versions of the GPQ showed excellent psychometric properties in Colombian samples and a one-factor structure. The two GPQ versions are provided in the Electronic Supplemental Material (ESM 1) in both Spanish and English.

Depression, Anxiety, and Stress Scales - 21 (DASS-21; Lovibond & Lovibond, 1995; Spanish version by Ruiz, García-Martín, Suárez-Falcón, & Odriozola-González, 2017). The DASS-21 consists of 21 items that are responded to on a 4-point Likert-type scale (0 = *did not apply to me at all*; 3 = *applied to me very much, or most of the time*). It evaluates the emotional symptoms experienced during the last week. The Spanish version has shown adequate internal consistency, convergent construct validity, treatment sensitivity, and a hierarchical structure with a general factor and three second-order factors: Depression, Anxiety, and Stress (Ruiz, García-Martín, et al., 2017).

Acceptance and Action Questionnaire – II (AAQ-II; Bond et al., 2011; Spanish version by Ruiz, Langer, Luciano, Cangas, & Beltrán, 2013). The AAQ-II is a general measure of experiential avoidance. It consists of a 7-item, 7-point Likert-type scale (1 = *never true*; 7 = *always true*). The Spanish translation showed good internal consistency, a one-factor structure, treatment sensitivity, and convergent construct validity.

Cognitive Fusion Questionnaire (CFQ; Gillanders et al., 2014; Spanish version by Ruiz, Suárez-Falcón, Riaño-Hernández, & Gillanders, 2017). The CFQ is a measure of cognitive fusion. This instrument is a 7-item, 7-point Likert-type scale (1 = *never true*; 7 = *always true*). The Spanish translation has shown excellent internal consistency, a one-factor structure, measurement invariance across gender and clinical and nonclinical samples, treatment sensitivity, and convergent construct validity.

Valuing Questionnaire (VQ; Smout, Davies, Burns, & Christie, 2014; Spanish version by Ruiz, Suárez-Falcón, & Gil-Luciano, under review). The VQ is a self-report instrument that assesses the extent to which the individual lives consistently with his or her values. This instrument is a 10-item, 7-point Likert (0 = *not at all true*; 6 = *completely true*). The items are divided into two subscales: Progress and Obstruction. The Spanish version has shown good internal consistency, a two-factor structure, measurement invariance across gender and clinical and nonclinical samples, treatment sensitivity, and convergent construct validity.

Satisfaction with Life Survey (SWLS; Diener, Emmons, Larsen, & Griffin, 1985; Spanish version by Atienza, Pons, Balaguer, & García-Merita, 2000). The SWLS is a measure of self-perceived well-being. This instrument consists of a 5-item, 7-point Likert-type scale (1 = *strongly disagree*; 7= *strongly agree*). The SWLS has shown good internal consistency, a one-factor structure, and convergent construct validity.

Procedure

The Institutional Ethics Committee approved the procedures of this study. Participants responded to an anonymous internet survey distributed through the Internet and social media. All participants provided informed consent. After finishing data collection, a global report was sent to the participants who provided an email address for that purpose. No incentives were provided for participation.

Data analysis

No missing values were found in the dataset (see ESM 2). A robust diagonally weighted least squares (Robust DWLS) estimation method using polychoric correlations was used to conduct the confirmatory factor analyses (CFA). We computed the Satorra-Bentler chi-square test and the following goodness-of-fit indexes for the one for the GPQ-18 and GPQ-9: (a) the root mean square error of approximation (RMSEA), (b) the

comparative fit index (CFI), the non-normed fit index (NNFI), and (d) the standardised root mean square residual (SRMR). According to Hu and Bentler (1999), CFI and NNFI values above .90 indicate acceptable-fitting models, and above .95 represent a good fit to the data. Regarding SRMR, values of 0.08 represent a good fit, and values below 0.05 represent a very good fit to the data. Lastly, regarding the RMSEA, it has been suggested that values above 0.10 represent a poor fit, values between 0.08 and 0.10 a marginal fit, values below 0.08 represent an acceptable fit, and values below 0.05 constitute a good fit (Browne & Cudeck, 1992, 1993; Hooper, Coughlan, & Mullen, 2008).

It is important to note that the previous cutoffs were developed for analysing continuous data using the normal-theory maximum likelihood estimation method. Consequently, these cutoffs do not provide absolute thresholds for other estimation methods such as the Robust DWLS (MacCallum, Browne, & Sugawara, 1996). Conversely, these guidelines should be taken flexibly and globally, taking into account several goodness-of-fit indexes (Brown, 2015). For instance, according to Marsh, Hau, and Wen (2004), it is not possible to designate specific cutoffs for each fit index because, in addition to model fit, they are a function of sample size, the type of estimator used, and possible violations of distributional assumptions. Furthermore, Barrett (2007) indicates it is necessary to explore if there are options to improve the model, why these options are or are not adopted, and what are the consequences of considering this model as the definitive one.

Metric and scalar invariances across gender were tested following the recommendations by Jöreskog (2005), and Millsap and Yun-Tein (2004). Metric invariance requires that item factor loadings are invariant across men and women, whereas scalar invariance requires that item intercepts are also invariant. In doing so,

we compared the relative fit of three increasingly restrictive nested models: the multiple-group baseline model (it allowed the unstandardised factor loadings to vary across groups), the metric invariance model (it placed equality of factor loadings across groups), and the scalar invariance model (it placed equality in both the factor loadings and the item intercepts across groups). According to Chen (2007), the more constrained model was selected if the following criteria were met: (a) the difference in RMSEA (ΔRMSEA) was lower than .01; (b) the differences in CFI (ΔCFI) and NNFI (ΔNNFI) were equal to or greater than -.01.

To explore the internal consistency of the GPQ, Cronbach's alpha was computed, providing 95% confidence intervals (CI) with SPSS 25©. Descriptive data were also calculated. Last, Pearson correlations between the GPQ versions and other scales were calculated to assess convergent construct validity. Correlations were interpreted according to the guidelines provided by Lenhard and Lenhard (2016): small correlation between .10 and .20, medium between .21 and .36, and strong correlations >.36.

Results

Validity evidence based on internal structure

Dimensionality. Table 1 shows the results of the CFAs conducted for each GPQ version (ESM 3-4). The fit of the one-factor model of the GPQ-18 was good according to the values of CFI (.97), NNFI (.97), and SRMR (0.053). However, the fit was marginal according to the value of the RMSEA (0.10). Modification indices recommended allowing error terms between Items 1-2, 2-4, and 8-16 to correlate. When doing so, the fit of the model improved (RMSEA = 0.084, 90% CI [0.079, 0.089], CFI = .98, NNFI = .98, SRMR = 0.047). Overall, the fit of the one-factor model of the GPQ-18 can be considered as acceptable.

The fit of the one-factor model of the GPQ-9 was good according to the CFI (.96), NNFI (.95), and acceptable according to the SRMR (0.063). However, the fit was poor, according to the RMSEA (0.146). Following the recommendations of the modification indices did not significantly improve the RMSEA. Overall, the fit of the one-factor model of the GPQ-9 was not acceptable. Accordingly, we only computed the following analyses with the GPQ-18.

INSERT TABLE 1 ABOUT HERE

Measurement invariance. Table 2 shows the results of the metric and scalar invariance analyses with the GPQ-18 (see ESM 5-7). Parameter invariance was supported at both levels across gender because changes in RMSEA, CFI, and NNFI were lower than .01. Both the metric and the scalar invariance are met with the three criteria, so it can be assumed that the GPQ-18 showed measurement invariance across samples and gender.

INSERT TABLE 2 ABOUT HERE

Internal consistency, corrected item-total correlations, and GPQ-18 scores. The GPQ-18 showed excellent internal consistency (α = .95, 95% CI [.95, .96]). Corrected item-total correlations were high for all items in the GPQ-18 (from .51 for Item 17 to .85 for Item 13). The mean score of the sample on the GPQ-18 was 58.04 (*SD* = 20.21). There were no statistically significant difference in scores across gender (men: *M* = 57.12, *SD* = 19.47; women: *M* = 58.52, *SD* = 20.27; *t*(811) = -0.82, *p* = .41).

Convergent construct validity. Table 3 shows that the scores on the GPQ-18 showed strong positive correlations with experiential avoidance, cognitive fusion, and obstruction in values. Conversely, the correlations with progress in values and life satisfaction were negative and strong. Lastly, the GPQ-18 showed positive, moderate to strong, correlations with emotional symptoms.

INSERT TABLE 3 ABOUT HERE

Discussion

This study aimed to analyse the factor structure, measurement invariance across gender, internal consistency, and convergent construct validity of the GPQ in a Spanish sample. Overall, the study showed that the GPQ-18 seems to be a valid measure of generalised pliance in Spanish samples. Specifically, the GPQ-18 showed a one-factor structure, measurement invariance across gender, excellent internal consistency, and convergent construct validity. However, the short form of the GPQ (i.e., GPQ-9) showed a poor fit in the CFA due to high RMSEA values. Accordingly, we recommend using the complete GPQ (i.e., GPQ-18) with Spanish samples.

The findings of the current study add evidence regarding the factor structure and psychometric properties of the GPQ-18. Specifically, as in the original study with Colombian samples (Ruiz et al., 2019), the GPQ-18 has shown excellent internal consistency. Regarding factor structure, the fit of the one-factor model was somewhat worse than the one found in Ruiz et al. due to the high RMSEA value. However, note that the fit improved after allowing to correlate the error terms of three pairs of items. Further studies might consider analysing this model with correlated error terms in other Spanish-speaking samples. Lastly, the correlations found in this study with other measures are coherent with the ones found in previous studies (Ruiz et al., 2019; Waldeck et al., 2019). Specifically, this study contributes additional evidence of the positive correlations of the GPQ-18 with measures of experiential avoidance, cognitive fusion, values obstruction, and emotional symptoms, and the negative correlations with values progress and life satisfaction.

A unique contribution of this study was the finding of metric and scalar invariance of the GPQ-18 across gender. This is an important contribution because

establishing measurement invariance is needed to compare the scores across groups of participants (e.g., Greiff & Scherer, 2018). A previous study with the GPQ-C (Salazar et al., 2018) demonstrated measurement invariance across gender and group age (i.e., children and adolescents). These authors found that girls scored higher on the GPQ-C than boys. Conversely, the current study did not find differences in the GPQ-18 scores across gender. This might indicate that the levels of generalised pliance might equalise during adulthood. However, this suggestion needs to be taken with caution because the differences might be due to cultural differences between Spain and Colombia.

The results of this study have shown that the fit of the one-factor model of the GPQ-9 is questionable in Spain. This contrasts with the previous analyses by Ruiz et al. (2019) and Waldeck et al. (2019), in which the one-factor model of the GPQ-9 showed a good fit. This divergence in results might be due to the methodological deficiencies usually found in the development of short forms (Smith, McCarthy, & Anderson, 2000).

Some limitations of the study are worth mentioning. Firstly, we analysed the functioning of the GPQ in a convenience, online sample. Secondly, women were more represented in the sample. Thirdly, we did not analyse the psychometric properties of the GPQ in a clinical sample, as in Ruiz et al. (2019). Fourthly, the GPQ was only correlated with other self-report measures. Accordingly, further studies might correlate the GPQ with behavioural measures such as the WCST. Lastly, the mean scores on the GPQ-18 found in this study are higher than those found in similar samples in Colombia. Further studies should analyse measurement invariance of the GPQ across samples of these countries and compare their mean scores to analyse the differences in generalised pliance across these cultures. In this sense, note that a previous study has shown that Irish adolescents show higher scores on the GPQ-C than Colombian adolescents (Stapleton, Ruiz, & McHugh, in press).

In conclusion, this study showed preliminary data on the validity of the GPQ-18 in Spanish samples. Further research should analyse the criterion validity of the GPQ in Spanish samples as, for example, in O'Connor et al. (2019), and analyse measurement invariance of the GPQ across different cultures.

Electronic Supplementary Material

ESM 1. GPQ Versions (.docx). Spanish and English versions of the GPQ.

ESM 2. Dataset (.xlsx). Dataset used in confirmatory factor analyses.

ESM 3. GPQ18 (.out). LISREL OUT files containing syntax and results of the CFA of the GPQ-18.

ESM 4. GPQ9 (.out). LISREL OUT files containing syntax and results of the CFA of the GPQ-9.

ESM 5. GPQGenderBL (.out). LISREL OUT files containing syntax and results of the MG baseline model.

ESM 6. GPQGenderMetric (.out). LISREL OUT files containing syntax and results of the metric invariance model.

ESM 7. GPQGenderScalar (.out). LISREL OUT files containing syntax and results of the scalar invariance model.

References

Atienza, F. L., Pons, D., Balaguer, I., & García-Merita, M. (2000). Propiedades psicométricas de la Escala de Satisfacción con la Vida en adolescentes [Psychometric properties of the Satisfaction with Life Scale in adolescents]. *Psicothema, 12*, 314-319.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*, 815-824. https://doi.org/10.1016/j.paid.2006.09.018

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., … Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire – II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*, 676-688. http://dx.doi.org/10.1016/j.beth.2011.03.007

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: Guilford.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230-258. https://doi.org/10.1177/0049124192021002005

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 464-504. https://doi.org/10.1080/10705510701301834

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment, 49*, 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Gillanders, D. T., Bolderston, H., Bond, F. W., Dempster, M., Flaxman, P. E., Campbell, L., … Remington, B. (2014). The development and initial validation of the Cognitive Fusion Questionnaire. *Behavior Therapy, 45*, 83-101. http://dx.doi.org/10.1016/j.beth.2013.09.001

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 38*, 404-411. https://doi.org/10.1037/h0059831

Greiff, S., & Scherer, R. (2018). Still comparing apples with oranges? Some thoughts on the principles and practices of measurement invariance testing. *European Journal of Psychological Assessment, 34*, 141-144. https://doi.org/10.1027/1015-5759/a000487

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., McEnteggart, C., Gys, J., & Hasler, C. (in press). Exploring the potential impact of relational coherence on persistent rule-following: The first study. *Learning & Behavior*, 1-19. https://doi.org/10.3758/s13420-019-00399-0

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Plenum Press.

Hayes, S. C., & Hayes, L. J. (1989). The verbal action of the listener as a basis for rule-governance. In S. C. Hayes (Ed.), *Rule-Governed Behavior* (pp. 153-190). New York: Plenum Press. https://doi.org/10.1007/978-1-4757-0447-1_5

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York: Guilford Press.

Hayes, S. C., Zettle, R., & Rosenfarb, I. (1989). Rule-following. In S.C. Hayes (Ed.), *Rule governed behavior* (pp. 191-220). New York: Plenum Press. https://doi.org/10.1007/978-1-4757-0447-1_6

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*, 53-60.

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. http://dx.doi.org/10.1080/10705519909540118

Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL*. Lincolnwood, IL. Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1999). *LISREL 8.30*. Chicago, IL: Scientific Software International.

Kissi, A., Hughes, S., Mertens, G., Barnes-Holmes, D., De Houwer, J., & Crombez, G. (2017). A systematic review of pliance, tracking, and augmenting. *Behavior Modification*, *41*, 683-707. https://doi.org/10.1177/0145445517693811

Lenhard, W., & Lenhard, A. (2016). *Calculation of Effect Sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*, 335-343. https://doi.org/10.1016/0005-7967(94)00075-U

Luciano, C., Valdivia-Salas, S., Cabello, F., & Hernández, M. (2009). Developing self-directed rules. In R. A. Rehfeldt, & Y. Barnes-Holmes (Eds.), *Derived relational responding: Applications for learners with autism and other developmental disabilities* (pp. 335-352). Oakland, CA: New Harbinger.

Luciano, C., Valdivia-Salas, S., & Ruiz, F. J. (2012). The self as the context for rule-governed behavior. In L. McHugh & I. Stewart (Eds.), *The self and perspective taking: Research and applications* (pp. 143-160). Oakland, CA: Context Press.

MacCallum, R. C., Browne, M. W., & Sugawara, H., M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-49.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341. https://doi.org/10.1207/s15328007sem1103_2

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*, 479-515. https://doi.org/10.1207/S15327906MBR3903_4

O'Connor, M., Byrne, P., Ruiz, F. J., & McHugh, L. (2019). Generalised pliance in relation to contingency insensitivity and mindfulness. *Mindfulness, 10*, 833-840. https://doi.org/10.1007/s12671-018-1046-5

Ruiz, F. J., García-Martín, M. B., Suárez-Falcón, J. C., & Odriozola-González, P. (2017). The hierarchical factor structure of the Spanish version of Depression Anxiety and Stress Scale-21. *International Journal of Psychology and Psychological Therapy, 17*, 97-105.

Ruiz, F. J., Langer, A. I., Luciano, C., Cangas, A. J., & Beltrán, I. (2013). Measuring experiential avoidance and psychological inflexibility: The Spanish translation of the Acceptance and Action Questionnaire. *Psicothema*, *25*, 123-129. https://doi.org/10.7334/psicothema2011.239

Ruiz, F. J., Peña-Vargas, A., Ramírez, E. S., Suárez-Falcón, J. C., García-Martín, M. B., García-Beltrán, D. M., … Sánchez, P. D. (in press). Efficacy of a 2-session RNT-focused ACT protocol for depression and generalized anxiety disorder: A randomized waitlist control trial. *Psychotherapy*. http://dx.doi.org/10.1037/pst0000273

Ruiz, F. J., Suárez-Falcón, J. C., Barbero-Rubio, A., & Flórez, C. L. (2019). Development and initial validation of the Generalised Pliance Questionnaire. *Journal of Contextual Behavioral Science*, *12*, 189-198. https://doi.org/10.1016/j.jcbs.2018.03.003

Ruiz, F. J., Suárez-Falcón, J. C., Riaño-Hernández, D., & Gillanders, D. (2017). Psychometric properties of the Cognitive Fusion Questionnaire in Colombia. *Revista Latinoamericana de Psicología*, *49*, 80-87. http://dx.doi.org/10.1016/j.rlp.2016.09.006

Salazar, D. M., Ruiz, F. J., Flórez, C. L., & Suárez-Falcón, J. C. (2018). Psychometric Properties of the Generalised Pliance Questionnaire-Children. *International Journal of Psychology and Psychological Therapy*, *18*, 273-287.

Sierra, M. A., Ruiz, F. J., Flórez, C. L., Riaño-Hernández, D., & Luciano, C. (2016). The role of common physical properties and augmental functions in metaphor effect. *International Journal of Psychology and Psychological Therapy*, *16*, 265-279.

Skinner, B. F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.), *Problem solving: Research, method and theory* (pp. 133-171). New York: John Wiley & Sons.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological assessment, 12*, 102-111. https://doi.org/10.1037/1040-3590.12.1.102

Smout, M., Davies, M., Burns, N., & Christie, A. (2014). Development of the Valuing Questionnaire (VQ). *Journal of Contextual Behavioral Science, 3*, 164-172. https://doi.org/10.1016/j.jcbs.2014.06.001

Stapleton, A., & McHugh, L. (2020). Adolescent correlates of the Generalised Pliance Questionnaire–Children. *Journal of Contextual Behavioral Science, 15*, 131-134.

Stapleton, A., Ruiz, F. J., & McHugh, L. (in press). Comparative investigation of adolescents' generalised pliance and psychological inflexibility across cultural contexts. *The Psychological Record*.

Törneke, N., Luciano, C., & Valdivia-Salas, S. (2008). Rule-governed behavior and psychological problems. *International Journal of Psychology and Psychological Therapy, 8*, 141-156.

Waldeck, D., Pancani, L., & Tyndall, I. (2019). An examination of the construct validity of the Generalised Pliance Questionnaire. *Journal of Contextual Behavioral Science, 11*, 50-54. https://doi.org/10.1016/j.jcbs.2018.12.003

Zettle, R. D., & Hayes, S. C. (1982). Rule-governed behavior: A potential theoretical framework for cognitive-behavior therapy. *Advances in Cognitive-Behavioral Research and Therapy, 1*, 73-118.

Table 1

Goodness-of-Fit Indexes of the Two Versions of the GPQ

Goodness-of-fit indicators	GPQ-18	GPQ-9
RMSEA	0.100	0.146
[90% CI]	[0.095, 0.110]	[0.135, 0.157]
CFI	0.97	0.96
NNFI	0.97	0.95
SRMR	0.053	0.063
S-Bχ^2 (*df*)	1283.274(135)	513.040(27)

Note. CFI = Comparative Fit Index; GPQ = Generalised Pliance Questionnaire; NNFI = Non-Normed Fit Index; RMSEA = Root Mean Square Error of Approximation; *S-B*χ^2 = Satorra-Bentler Chi-Square Test; SRMR = Standardized Root Mean Square Residual.

Table 2

Metric and Scalar Invariance Across Gender in the GPQ-18

Model	RMSEA	ΔRMSEA	CFI	ΔCFI	NNFI	ΔNNFI
MG Baseline model	.1004		.971		.967	
Metric invariance	.0983	.0021	.971	.000	.969	.002
Scalar invariance	.0980	.0003	.969	-.002	.969	.000

Table 3

Pearson Correlations between the GPQ-18 and Other Relevant Self-Report Measures

	GPQ-18
AAQ-II – Experiential avoidance	.53*
CFQ – Cognitive fusion	.51*
VQ-Progress – Progress in values	-.37*
VQ-Obstruction – Obstruction in values	.49*
SWLS – Life satisfaction	-.36*
DASS-Total – Emotional symptoms	.42*
DASS-Depression	.42*
DASS-Anxiety	.33*
DASS-Stress	.39*

Note. *p < .001 AAQ-II = Acceptance and Action Questionnaire – II; CFQ = Cognitive Fusion Questionnaire; DASS = Depression, Anxiety, and Stress Scale; GPQ = Generalised Pliance Questionnaire; SWLS = Satisfaction with Life Scale; VQ = Valuing Questionnaire.