

FACTORIAL VALIDITY AND ITEM RESPONSE THEORY ANALYSIS OF THE GRATITUDE QUESTIONNAIRE IN AN ITALIAN SAMPLE

GIULIA FUOCHI
ALBERTO VOCI
CHIARA VENEZIANI
UNIVERSITY OF PADOVA

The aim of this paper is to assess the diversified behaviors of the Gratitude Questionnaire (GQ-6) items. After providing evidence on the monofactorial structure of the GQ-6 in the Italian context (pilot study), we examined the performance of the questionnaire (main study) relying on item response theory (IRT). Results prove that the sixth item produces a lower discrimination parameter and scarcer information compared to the other items, while the opposite happens for the first two items. The test information function shows that the GQ-6 better captures respondents' gratitude when they are in low to median positions of the trait, compared to higher values of the latent trait. Our findings suggest that the sixth item may be discarded from the GQ-6 in case of time constraints in the research design, and that there might be a response tendency toward the upper response options.

Key words: Gratitude; Gratitude Questionnaire; Confirmatory factor analysis; Item response theory; Validation.

Correspondence concerning this article should be addressed to Alberto Voci, Department FISPPA – Section of Applied Psychology, University of Padova, Via Venezia 14, 35131 Padova (PD), Italy. Email: alberto.voci@unipd.it

Sixteen years have passed since the publication of the Gratitude Questionnaire (McCullough, Emmons, & Tsang, 2002), and the construct of gratitude has never lost its popularity, being present in a large number of studies across different psychological disciplines. Gratitude has been defined as a positive emotion arising from the perception that other subjects, including in this set not only individuals but also non-human entities like nature or God, contributed to an experienced positive outcome (e.g., Emmons & McCullough, 2003; McCullough et al., 2002; McCullough, Kilpatrick, Emmons, & Larson, 2001).

Being aware of having received acts of benevolence (Emmons & McCullough, 2003) engenders a number of positive consequences: social bonds and social support are stimulated and reinforced (e.g., Algoe, Haidt, & Gable, 2008; Wood, Maltby, Gillett, Linley, & Joseph, 2008), the receiver is motivated to behave prosocially toward the benefactor and other people (Bartlett, Condon, Cruz, Baumann, & DeSteno, 2012; McCullough et al., 2001), and well-being is enhanced in many ways. These include reduced depression and stress (Lambert, Fincham, & Stillman, 2012; Wood et al., 2008), and increased positive affect, life satisfaction (e.g., Emmons & McCullough, 2003; Wood, Froh, & Geraghty, 2010), and psychological well-being (Wood et al., 2010).

In the studies cited previously, gratitude was either induced by experimental manipulations (Bartlett et al., 2012; Emmons & McCullough, 2003), or measured by naturally occurring gratitude acts (Algoe et al., 2008), or, most of the time, assessed by self-report measures of dispositional gratitude (e.g., Lambert

et al., 2012; Wood et al., 2008). The most frequently used scale of dispositional gratitude is the six-item Gratitude Questionnaire (GQ-6; McCullough et al., 2002).

THE GRATITUDE QUESTIONNAIRE

The six-item Gratitude Questionnaire was developed, starting from a set of 39 items, to capture overall frequency, intensity, span, and density of grateful affect in an individual (McCullough et al., 2002). These four characteristics respectively portray how many times, how much, for how many aspects of their life, and to how many people individuals feel grateful and simultaneously contribute to the detection of individual differences in dispositional gratitude (McCullough et al., 2002). Indeed, the GQ-6 has a monofactorial structure, and gratitude is assessed by its total score. The scale has been validated also in an adolescent English-speaking sample (Froh et al., 2011), and in Chilean Spanish (Langer, Ulloa, Aguilar-Parra, Araya-Véliz, & Brito, 2016), Dutch (Jans-Beken, Lataster, Leontjevas, & Jacobs, 2015), and Chinese in Taiwan (Chen, Chen, Kee, & Tsai, 2008) adult samples.

Interestingly, all these validations share a common pattern: the sixth item always produces a low, and sometimes very low, factor loading, for example equal to .21 in Froh et al. (2011), and smaller than .08 in Chen et al. (2008) and Langer et al. (2016). At the same time, the first two items show very high factor loadings, and the others place themselves in the middle. This pattern is present to a smaller extent also in other studies with English-speaking samples (see Table A1 in Froh et al., 2011), but without reaching the extremely low values observed in Chen et al. (2008) and Langer et al. (2016), which resulted in the exclusion of the sixth item from the scale. As Chen et al. (2008) commented that because Item 6 indicates a temporal definition of gratitude (“Long amounts of time can go by before I feel grateful to something or someone”), it could have a less linear behavior or it could be affected by the specific life experiences of the sample. Another possible explanation may be that the temporal reference of the sixth item requires an evaluation from the respondent on what is a “long amount of time,” and this might raise doubts on the interpretation of the item or generate a missed identification of the “truly grateful” person. Indeed, people with high dispositional gratitude might even answer positively (e.g., “agree” or “slightly agree”) to this item, depending on their personal perception of time and life experiences. The semantic distance between the sixth item and the others is also likely to decrease the correlations between the items.

Despite the heterogeneous behaviors of the six items composing the GQ-6, the poor performance of the sixth item, and the wide use of the scale, no studies until now have tried to clarify its characteristics and performance.

AIMS OF THE PAPER

In this paper, we intended to assess the diversified behaviors of the GQ-6 items, by applying — to our knowledge for the first time — item response theory (IRT) analyses on the GQ-6. In particular, we wish to explain the poor performance of Item 6, an issue which emerged in previous studies (Chen et al., 2008; Froh et al., 2011; Langer et al., 2016) and led to shorter versions of the scale, excluding this item (Chen et al., 2008; Langer et al., 2016).

In the pilot study ($N = 629$), we proved the factorial validity of the GQ-6 in the Italian context, replicating the unidimensional structure of the original validation study (McCullough et al., 2002), and

comparing the original 7-point response scale with the 5-point response scale used in the subsequent analyses. Other studies used a 5-point response scale for the GQ-6 (e.g. Chan, 2013; Chang, Lin, & Chen, 2012; Neto, 2007), but we decided to see if the results' pattern changed depending on the response scale, before running the main analysis. In the main study ($N = 2,496$), we examined in depth the GQ-6 items' characteristics and performance, relying on IRT analyses for polytomous items, in particular on the graded response model (Samejima, 1969, 1995). IRT analyses shed light on the ability of the single items to distinguish individuals by their latent trait, in our case gratitude, and on the overall performance of the questionnaire, that is the GQ-6, for each value of the latent trait.

PILOT STUDY

In this study, we explored the factorial structure of the GQ-6 in the Italian context, also comparing the original 7-point response scale with the 5-point response scale we used in the main study. We randomly split a sample selected from the general population into two subsamples and assigned to each of them a questionnaire including one of the two versions of the GQ-6 and sociodemographic questions. Thus, one subsample answered the 7-point GQ-6 and the other subsample answered the 5-point GQ-6. Then, we tested the one-factor structure found in previous studies (Chen et al., 2008; Froh et al., 2011; McCullough et al., 2002) with confirmatory factor analysis (CFA), separately on the two subsamples. We expected to reproduce the factorial structure of the GQ-6 already found in its validation article (McCullough et al., 2002), and to find similarity and consistency between the two alternative versions, allowing us to use the 5-point GQ-6 in subsequent analyses.

Method

Participants

The subsample that was assigned to the 7-point GQ-6 filled in 336 questionnaires (Subsample 1), while the other filled in 293 questionnaires (Subsample 2). In both subsamples, 0.01% of respondents left several GQ-6 items blank. Because of the small size of this percentage and the type of analysis we used (not allowing the management of missing values with full-information maximum likelihood), we discarded these respondents. Gender was balanced in both subsamples, with women representing 52% in Subsample 1 and 53% in Subsample 2. The two subsamples were very similar also in terms of age, $M_1 (SD_1) = 29.8 (12.8)$, $M_2 (SD_2) = 29.0 (12.5)$, and well-being, as measured by a single-item question on life satisfaction ranging from 1 to 10, $M_1 (SD_1) = 6.25 (2.12)$, $M_2 (SD_2) = 6.18 (2.11)$.

Procedure and Measures

As a first step, we translated the six items of the GQ-6 with the support of a bilingual psychologist who has spent half of her life in Australia and the other half in Italy, and currently lives in Italy. Following test translation guidelines (e.g., Gudmundsson, 2009), we then contacted a bilingual professional (native Italian speaker) to have a backward translation of our Italian version of the GQ-6. The backward translation

confirmed the appropriateness of the forward translation¹. The questionnaires, including the GQ-6, background questions, and other variables not reported here, were filled in online. Before answering the questionnaire, respondents were briefly informed about its general focus, the anonymity of their responses, and the possibility to withdraw from the research study at any time.

Gratitude Questionnaire (GQ-6). The Italian version of the 6-item Gratitude Questionnaire was employed. Response options were on a Likert-type scale from 1 (*strongly disagree*) to 7 (*strongly agree*) in Subsample 1 and from 1 (*strongly disagree*) to 5 (*strongly agree*) in Subsample 2. Before the analyses, the third and the sixth items were reverse-coded so that higher scores indicated higher gratitude. In the 7-point subsample, the average gratitude was 4.94 ($SD = 1.06$), whereas the average gratitude of the 5-point subsample was 3.62 ($SD = 0.67$). Cronbach's alphas were, respectively, .78 and .75.

Results and Discussion

Before performing the two parallel CFAs, we plotted the items. We noticed a severe asymmetry in most of the items, so we checked items' normality. In both subsamples, items' normality was rejected by both the Shapiro-Wilk test ($0.873 < W < 0.946$ in the 7-point version, $0.855 < W < 0.909$ in the 5-point version, $p < .001$, for all six items in both the subsamples) and Mardia's test. Indeed, Mardia's test yielded similar results in the 7-point (multivariate skewness: $b_{1p} = 4.11$, $p < .001$; multivariate kurtosis: $b_{2p} = 54.84$, $p < .001$) and in the 5-point subsample (multivariate skewness: $b_{1p} = 3.42$, $p < .001$; multivariate kurtosis: $b_{2p} = 53.44$, $p < .001$). Therefore, we treated the items as ordinal and performed all the subsequent CFAs with the diagonally weighted least squares (DWLS), an estimator specifically designed for ordinal items and Likert-scale data (Flora & Curran, 2004) and giving more accurate estimates than maximum likelihood (ML) methods in cases of nonnormality (Li, 2016a, 2016b).

We performed a CFA testing the same model advanced by the authors of the questionnaire (McCullough et al., 2002), that is, all the items loading on one factor measuring gratitude, separately in the two subsamples. For these analyses, we used the package Lavaan (Rosseel, 2012) in R (R Core Team, 2017). As thresholds for an acceptable fit, we considered a root mean square error of approximation (RMSEA) smaller or equal to .08, a comparative fit index (CFI) and Tucker-Lewis index (TLI) higher than .95, and a standardized root mean square residual (SRMR) smaller than .10 (Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Except for the RMSEA in the subsample with the 7-point response scale, fit indexes were below the thresholds (7-point GQ-6: CFI = .991; TLI = .984; RMSEA = .151; SRMR = .079; 5-point GQ-6: CFI = .994; TLI = .991; RMSEA = .084; SRMR = .058). In the version with the 7-point scale the factor loadings of the first five items ranged from .45 to .95, while the sixth item yielded a standardized loading equal to .14 (Table 1), similarly to Langer et al. (2016) and Chen et al. (2008). The version with the 5-point scale generated similar factor loadings, but Item 6 in this case produced a loading equal to .23 (Table 1).

As in Chen et al. (2008) and Langer et al. (2016), we tested also the monofactorial solution with the first five items (GQ-5). Fit indexes improved (7-point GQ-5: CFI = .999; TLI = .997; RMSEA = .074; SRMR = .034; 5-point GQ-5: CFI = .998; TLI = .996; RMSEA = .064; SRMR = .039). However, as the GQ-6 with the 5-point scale had good factorial validity, and because we aimed at studying the performance of the original GQ-6, in the main study we relied on the GQ-6 with the 5-point response scale.

TABLE 1
Items of the GQ-6 and standardized factor loadings
(Pilot study; 7-point and 5-point versions)

Item position	Factor loadings	
	7-point	5-point
1	.93	.93
2	.95	.90
3	.45	.41
4	.74	.68
5	.72	.61
6	.14	.23

Note. Items 3 and 6 were reverse-coded.

MAIN STUDY

The aim of this study was to analyze in depth the GQ-6 items' characteristics and performance, relying on IRT analyses. We assumed a continuous latent trait and we treated the 5-point response scale of the GQ-6 as ordinal. We performed a graded response model (Samejima, 1969, 1995), suitable for polytomous items, and derived the related item characteristic curves and item information functions. Based on previous validations and reported GQ-6 factor loadings (e.g., Chen et al., 2008; Froh et al., 2011), we expected the sixth item to provide less information compared to the others, and the first two items to show the best performance among the six.

Method

Participants

Data were collected from a convenience sample of 2,496 Italian adults from the general population, who filled in the GQ-6 across the last five years and completed at least four out of six items. Only four respondents were discarded because of partial completion, reaching the current size of 2,496 individuals. Participants' age ranged from 18 to 80 years ($M = 33.15$; $SD = 13.78$), and women represented 61.18% of the sample. Concerning educational level, 0.05% of the respondents had attended only primary school, 8.69 % had attended secondary school, 41.31% had a high school diploma, and 41.15% had a university degree, while 8.80% did not report their education. Students represented 40.26% of the sample; 10.86% of the sample were professionals, teachers in secondary schools or academics, 27.76% were retailers, employees or teachers in primary schools, 8.69% were manual or office workers, and 8.09% were unemployed, retired or homemakers. The remaining part of the sample did not indicate any occupation.

Procedure and Measures

The questionnaires, including the GQ-6, background questions, and other variables not reported in this study, were filled in online or on paper. In both cases, before answering the questionnaire respondents were briefly informed about its general focus, the anonymity of their responses, and the possibility to withdraw from the research study at any time.

Gratitude Questionnaire (GQ-6). The Italian version of the 6-item Gratitude Questionnaire with 5-point response scale was employed. Response options were on a Likert-type scale from 1 (*strongly disagree*) to 5 (*strongly agree*), and Cronbach's alpha was .78. Before all the analyses, the third and the sixth items were reverse-coded so that higher scores indicated higher gratitude. The reverse-coding of those two items is reminded with a capital "R" in the tables and figures. We also computed a dispositional gratitude aggregated score, by averaging the six items for each individual.

Results

First, we explored the distribution of dispositional gratitude individual scores along the GQ-6 response scale, that is, from 1 to 5. We created a frequency polygon (Figure 1) showing the percentage of subjects located in each point of the response scale. Figure 1 shows that most of the sample is located in the upper half of the response scale, indicating a negatively skewed distribution, and the prevalence of high dispositional gratitude scores. Then, we moved to the IRT part of the analysis.

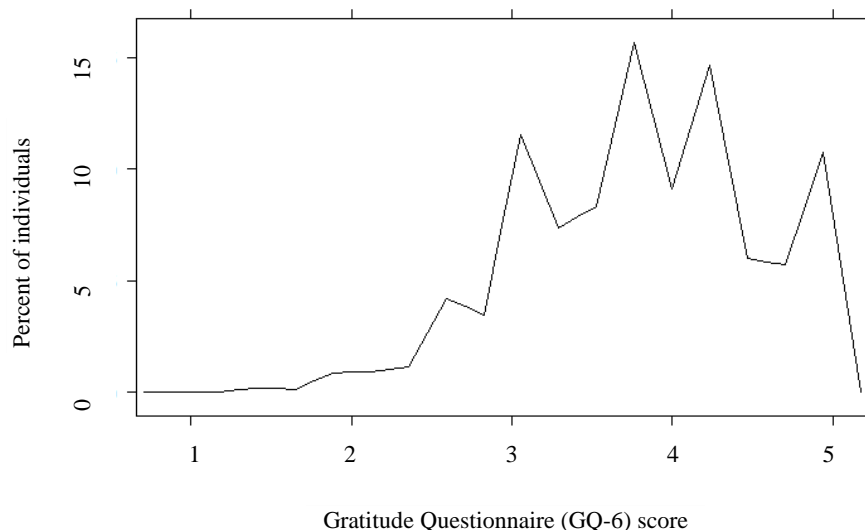


FIGURE 1
Frequency polygon showing percentage of individuals along the GQ-6 response scale.

The two basic assumptions of latent trait models are interconnected: the latent structure has to be unidimensional, and the items have to be locally independent of each other, conditionally on the latent trait, which is assumed continuous and depicted by categorical items. Both the previous literature (e.g., Chen et al., 2008; Froh et al., 2011; McCullough et al., 2002) and the pilot study have showed that the GQ-6 has a

unidimensional structure, depicting the underlying trait of gratitude. We performed a graded response model (GRM; Samejima, 1969, 1995) on the GQ-6. The GRM is a family of models dealing with ordered polytomous response categories, such as Likert scales (Samejima, 1995). It models the probability of scoring in a certain response category or higher, given the underlying trait (Samejima, 1969) and it computes thresholds between each response option and the ordinaly higher ones, relying on cumulative response distributions for each item (Samejima, 1969). We performed an unconstrained graded response model, that is, without imposing equal discrimination parameters across items, and we reported the coefficients' estimates under the usual IRT parameterization, namely with discrimination and threshold parameters. All the IRT analyses were performed with the package ltm (Rizopoulos, 2006) in R (R Core Team, 2016).

IRT Results

Results are reported in Table 2. Discrimination parameters portray the items' ability to differentiate among individuals who are high or low on the latent trait. Larger discrimination parameters mean that the likelihood of endorsing an item increases more rapidly as the underlying trait increases. According to Baker (2001), the differentiation ability is low when the discrimination parameter is less than 0.64, moderate when the parameter is in the range 0.65-1.34, high for the range 1.35-1.69, and very high for values of 1.70 and over. Item threshold parameters (difficulty parameters) indicate the point along the latent trait where the probability of selecting a given response option is 50%, namely the crossover point between two response categories. The number of item threshold parameters is determined by the number of response categories minus one, in our case four. Table 2 shows that, except for Item 6, GQ-6 items have a moderate to very high ability to differentiate among trait levels, and the first two items of the GQ-6 perform best in terms of discrimination, as we expected.

TABLE 2
Item discrimination and threshold parameters for the GQ-6 (main study)

Item	Discrimination	Threshold 1-2	Threshold 2-3	Threshold 3-4	Threshold 4-5
1	4.621	-2.519	-1.720	-0.875	0.101
2	2.985	-2.324	-1.326	-0.317	0.655
3R	1.416	-3.290	-2.010	-1.073	0.304
4	1.481	-2.769	-1.334	-0.152	1.342
5	1.268	-3.449	-2.137	-0.954	0.791
6R	0.676	-4.795	-2.227	-0.384	1.920

Note. Items 3R and 6R are the reverse-coded versions of Items 3 and 6.

We computed individual parameter estimates, that is, the individuals' estimated abilities, through factor scores from this model, and we compared them with item parameter estimates, that is, the thresholds. The resulting person-item map is reported in Figure 2, showing that GQ-6 items are targeted mainly at people with low to median scores of the latent trait, who constitute around half of the sample. The only exception is Item 4, which extends more uniformly than the other items along the latent trait.

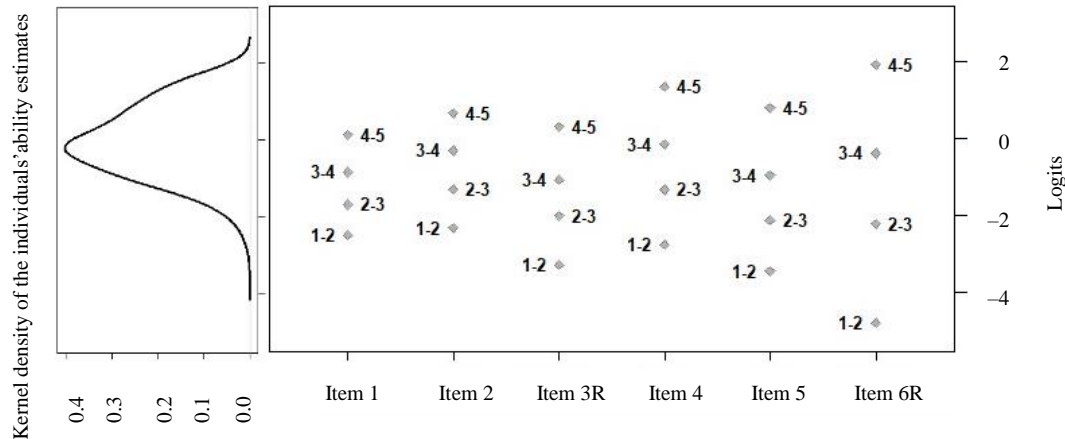


FIGURE 2
Person-item map comparing item and person parameter estimates.

Figure 3 allows a visual inspection of the items' performance for each response category from one to five, as reported in the legend in the top right side of each graph. The curves indicate the probability (on the y-axis) of choosing a certain response option (from one to five) for each value of the latent trait. The steeper curves of the first two items confirm our hypotheses on their discriminative power, while the flatter curves of the sixth item sanction both our expectations and previous research results finding a small factor loading for this item (e.g., Froh et al., 2011) or an acceptable model fit only by deleting it (Chen et al., 2008). In fact, for Items 1 and 2 the probability of choosing the highest response option (indicating high gratitude) is approximately equal to one in people with high values of the latent trait, and the exactly opposite situation can be observed for low values of the latent trait. Conversely, for the sixth item there is not much difference in the probability of choosing one response option over the others, along the latent trait continuum. Moreover, Figure 3 shows that the middle category (the “*neither agree nor disagree*” option) is usually chosen by individuals who are in the lower rather than in the median positions of the latent trait, suggesting that there might be a tendency, or even a “bias,” toward the upper response options. This is consistent with the thresholds reported in Table 2.

We then plotted in Figure 4 each item's information, namely the accuracy of each item in measuring any value of the latent trait (e.g., de Ayala, 2009; Samejima, 1995), and in Figure 5 the total information of the Gratitude Questionnaire, computed as the sum of items' information. As information functions depend on the discrimination and threshold parameters of the items, they are consistent with what already observed in Table 2 and Figure 3: the sixth item is the one with the weakest contribution to the total information, whereas Items 1, 2, and 4 contribute in an important way, although with differences between the three. Indeed, Item 4 and to some extent Item 3 provide some information also for higher values of the latent trait, outperforming Items 1 and 2 in those positions. The peaks in the information curves of Items 1 and 2 suggest that these items measure gratitude with unequal precision along the trait scale (Baker, 2001), with maximum inaccuracy in the high values of the trait. Flat — and elevated — information curves would indicate that the items are accurate in measuring gratitude for larger ranges of the latent trait, that is, for people with different levels of underlying dispositional gratitude.

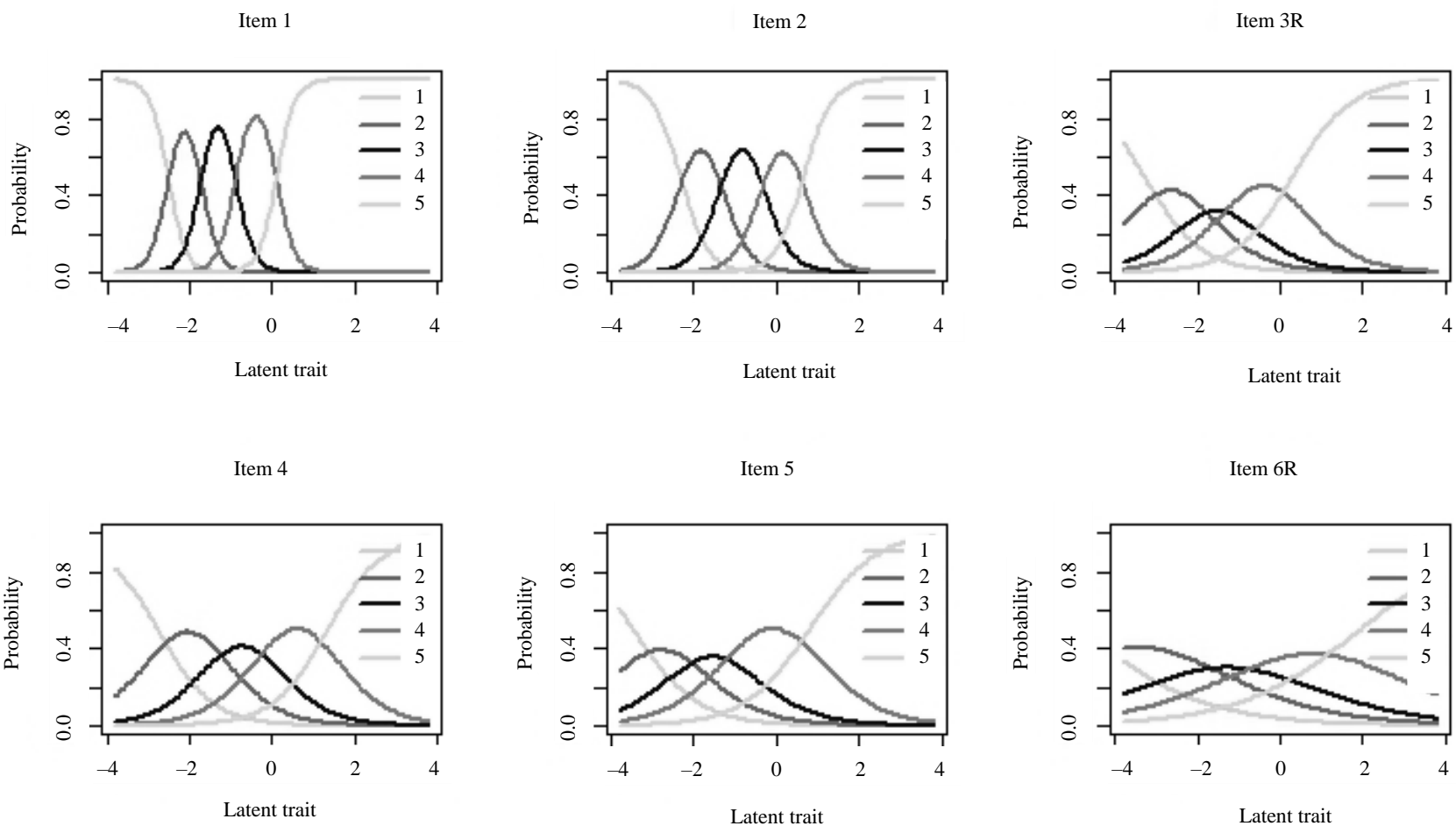


FIGURE 3
Item response category characteristic curves for the GQ-6.

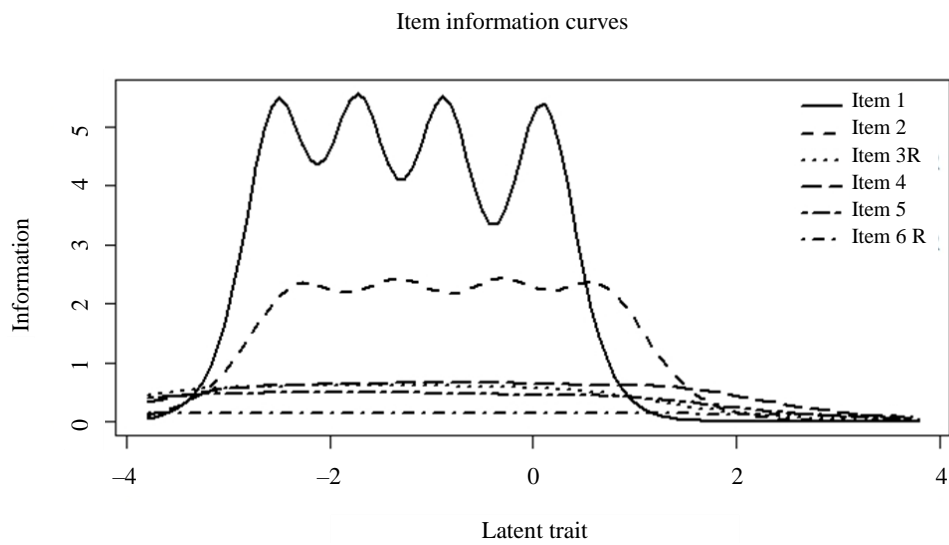


FIGURE 4
Item information curves for each item of the GQ-6.

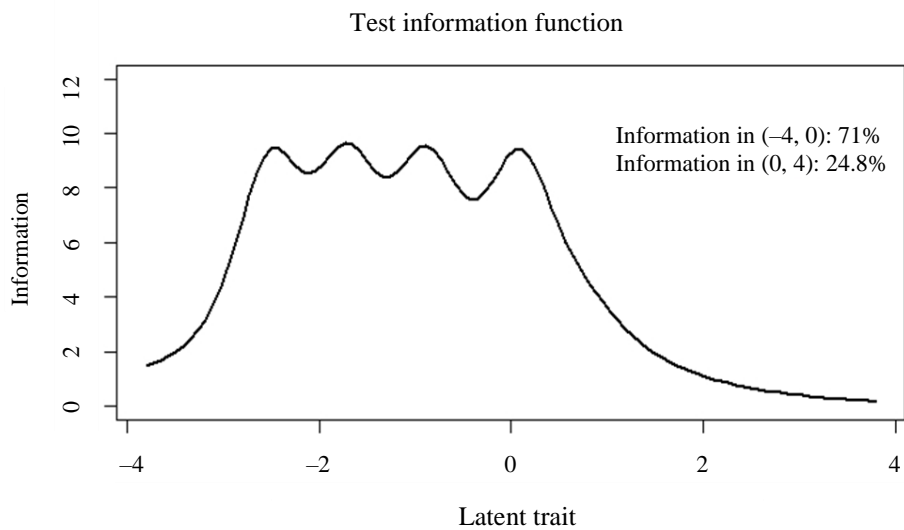


FIGURE 5
Global test information curve for the GQ-6.

Likewise, the total information of a questionnaire describes how well the test is doing in estimating the latent trait over the whole range of trait scores (e.g., Baker, 2001; Rizopoulos, 2006). Figure 5 shows that the GQ-6 does not provide information uniformly along the latent trait. Respondents' gratitude is best captured when they are in low to median positions of the trait, while gratitude is estimated less accurately in people with moderate to very high values of the latent trait (Baker, 2001). Overall, results suggest that it is quite unlikely that individuals filling in the GQ-6 choose the lowest response options, and the ones

who do this are on the lowest positions of the gratitude latent trait. One possible explanation could be linked to socially desirable responding, which is usually positively, but weakly, correlated with GQ-6 scores (McCullough et al., 2002).

GENERAL DISCUSSION

In this paper, we explored the performance of the most frequently used measure of dispositional gratitude, the Gratitude Questionnaire. First, we confirmed the unidimensional factorial structure of the GQ-6 in its Italian version, and then we shed light on the diversified behaviors of the items of the scale. Considering both CFAs and IRT analyses, items 1 and 2 showed the highest factor loadings, discrimination parameters, and amounts of information. Conversely, Item 6 produced low or very low factor loadings, and consistently showed low discrimination and scarce information. Factor loadings were in line with previous research (see Froh et al., 2011), and together with IRT results confirmed the issues already raised by Chen et al. (2008) on the sixth item of the GQ-6, that is, its uncertain relevance and limited consistency with the scale. As already hypothesized, the wording of the item, involving a temporal reference, might hinder the respondent's interpretation of it. One solution could be eliminating the sixth item from the GQ-6, as already done in previous studies (Chen et al., 2008; Langer et al., 2016), justifying this deletion not exclusively with the low factor loading, but especially with the very low amount of information produced by this item. Indeed, the information curve of the sixth item was almost flat for all the values of the latent trait, and it never went above the information curves of the other items. This means that there is not any position along the gratitude latent trait where the sixth item has an incremental utility compared to the others. Summarizing, Item 6 gives a very limited contribution to the overall information provided by the questionnaire. As with the 5-point response scale Item 6 showed a factor loading equal to .23, so it might be convenient to keep it to preserve the original structure of the scale whenever researchers are not faced with time constraints and other limitations due to the research design.

The questionnaire also showed an unexpectedly nonuniform behavior along the latent trait. Ideally, it might be desirable to have relatively flat information curves, so that the items and the global test are constantly accurate for almost any value of the latent trait (e.g., Baker, 2001). This does not happen with the GQ-6, where the items providing more information (items 1 and 2) are accurate especially for low values of the latent trait, and this affects the overall test information function. For future research on dispositional gratitude, it might be important to consider that the GQ-6 scores may be more able to describe the range of people with low to median latent gratitude, than very grateful people. Moreover, future studies could explore the response bias toward the upper categories of the GQ-6 response scale, and investigate the possible ways to measure more accurately the gratitude of individuals with fairly high or very high dispositional gratitude. This issue could be investigated for scales assessing other dispositions as well. It might be possible that also other self-report measures used to capture positive individual characteristics suffer the same problem.

NOTE

1. Italian version of the items is available upon request from the first author.

REFERENCES

- Algoe, S. B., Haidt, J., & Gable, S. L. (2008). Beyond reciprocity: Gratitude and relationships in everyday life. *Emotion*, 8, 425-429.
doi:10.1037/1528-3542.8.3.425
- Baker, F. (2001). The basics of item response theory. ERIC clearinghouse on assessment and evaluation. College Park, MD: University of Maryland.
- Bartlett, M. Y., Condon, P., Cruz, J., Baumann, J., & DeSteno, D. (2012). Gratitude: Prompting behaviours that build relationships. *Cognition & Emotion*, 26, 2-13.
doi:10.1080/02699931.2011.561297
- Chan, D. W. (2013). Subjective well-being of Hong Kong Chinese teachers: The contribution of gratitude, forgiveness, and the orientations to happiness. *Teaching and Teacher Education*, 32, 22-30.
doi:10.1016/j.tate.2012.12.005
- Chang, Y.-P., Lin, Y.-C., & Chen, L. H. (2012). Pay it forward: Gratitude in social networks. *Journal of Happiness Studies*, 13, 761-781.
doi:10.1007/s10902-011-9289-z
- Chen, L. H., Chen, M.-Y., Kee, Y. H., & Tsai, Y.-M. (2008). Validation of the Gratitude Questionnaire (GQ) in Taiwanese undergraduate students. *Journal of Happiness Studies*, 10, 655-664.
doi:10.1007/s10902-008-9112-7
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology*, 84, 377-389.
doi:10.1037/0022-3514.84.2.377
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 466-491.
doi:10.1037/1082-989X.9.4.466
- Froh, J. J., Fan, J., Emmons, R. A., Bono, G., Huebner, E. S., & Watkins, P. (2011). Measuring gratitude in youth: Assessing the psychometric properties of adult gratitude scales in children and adolescents. *Psychological Assessment*, 23, 311-324.
doi:10.1037/a0021590
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, 61, 29-45.
doi:10.1027/1901-2276.61.2.29
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
doi:10.1080/10705519909540118
- Jans-Beken, L., Lataster, J., Leontjevas, R., & Jacobs, N. (2015). Measuring gratitude: A comparative validation of the Dutch Gratitude Questionnaire (GQ6) and Short Gratitude, Resentment, and Appreciation Test (SGRAT). *Psychologica Belgica*, 55, 19-31.
doi:10.5334/pb.bd
- Lambert, N. M., Fincham, F. D., & Stillman, T. F. (2012). Gratitude and depressive symptoms: The role of positive reframing and positive emotion. *Cognition & Emotion*, 26, 615-633.
doi:10.1080/02699931.2011.595393
- Langer, A. I., Ulloa, V. G., Aguilar-Parra, J. M., Araya-Véliz, C., & Brito, G. (2016). Validation of a Spanish translation of the Gratitude Questionnaire (GQ-6) with a Chilean sample of adults and high schoolers. *Health & Quality of Life Outcomes*, 14, 1-9.
doi:10.1186/s12955-016-0450-6
- Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48, 936-949.
doi:10.3758/s13428-015-0619-7
-

- Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, 21, 369-387.
doi:10.1037/met0000093
- McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, 82, 112-127.
doi:10.1037/0022-3514.82.1.112
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, 127, 249-266.
doi:10.1037//0033-2909.127.2.249
- Neto, F. (2007). Forgiveness, personality and gratitude. *Personality and Individual Differences*, 43, 2313-2323.
doi:10.1016/j.paid.2007.07.010
- R Core Team. (2017). *R: A language and environment for statistical computing. R foundation for statistical computing*.
Retrieved from <https://www.R-project.org>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25.
doi:10.18637/jss.v017.i05
- Rosseel, Y. (2012). Lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36.
doi:10.18637/jss.v048.i02
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34, 1-97.
doi:10.1007/BF03372160
- Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, 60, 549-572.
doi:10.1007/BF02294328
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and goodness-of-fit models. *Methods of Psychological Research Online*, 8, 23-74.
- Wood, A. M., Froh, J. J., & Geraghty, A.W. A. (2010). Gratitude and well-being: A review and theoretical integration. *Clinical Psychology Review*, 30, 890-905.
doi:10.1016/j.cpr.2010.03.005
- Wood, A. M., Maltby, J., Gillett, R., Linley, P. A., & Joseph, S. (2008). The role of gratitude in the development of social support, stress, and depression: Two longitudinal studies. *Journal of Research in Personality*, 42, 854-871.
doi:10.1016/j.jrp.2007.11.003