Classical Perspectives of Controlling Acquiescence with Balanced Scales

Ricardo Primi, Nelson Hauck-Filho, Felipe Valentini, and Daniel Santos

Abstract Acquiescence, the tendency to agree regardless of the content of an item, is a commonly observed response style that may distort respondent scores. In the current study, we: (a) revised basic concepts of methods for measuring and controlling acquiescence, (b) describe some important properties of balanced scales, (c) examine if methods of controlling acquiescence provide ipsative scales, (d) explain the mechanism underlying the correction of acquiescence, and (e) compare the centering and standardizing correction methods. By using simulated data, we demonstrate that balanced scales are automatically controlled for acquiescence and that the scoring process does not yield ipsative scales. By contrast, the standardizing method of correction in fact undo the correction that takes place when using the centering method.

Keywords Acquiescence bias · Balanced scales · True-keyed items false keyed items

1 Introduction

1.1 What Is Acquiescence?

Acquiescence is the tendency to endorse the highest Likert categories regardless of the content of the item. One way to examine acquiescence is to include positively-

R. Primi (✉)
Universidade São Francisco, São Paulo, Brazil

EduLab21, Ayrton Senna Institute, São Paulo, Brazil

D. Santos
Universidade de São Paulo, São Paulo, Brazil

EduLab21, Ayrton Senna Institute, São Paulo, Brazil

N. Hauck-Filho · F. Valentini
Universidade de São Paulo, São Paulo, Brazil

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 322, https://doi.org/10.1007/978-3-030-43469-4_25

keyed (PK) and negatively-keyed items (NK), that is, markers of opposite poles of a trait. For instance, suppose an item designed to measure negative emotional regulation, such as $i+$: "*I adapt easily to new situations without worrying too much*," for which students must use the following scale to answer: "1" (not at all like me), "2" (little like me), "3" (moderately like me), "4" (a lot like me), and "5" (completely like me). Also, suppose an antonym paired item is included, such as $i-$: "*I have trouble controlling my anxiety in difficult situations*." A student very high in acquiescence will tend to endorse categories "4" or "5" of both items, which is semantically inconsistent. By contrast, a student who scores high in emotional regulation and low in acquiescence, that is, a person whose item responses are primarily driven by the trait content, will tend to give opposite responses, for example, "4" to $i+$ and "2" to $i-$.

Acquiescence represents a method factor, that is, a systematic source of variance unrelated to the target construct that researchers intend to measure (systematic error of measurement, McCrae 2015). It accounts for a sizable portion of the item's variance in questionnaires, especially when assessing children. It can distort the inter-item covariance matrix of an instrument (internal structure validity), and bias correlations with external variables (criterion validity, see: Primi et al. 2019a, b, c).

Some researchers propose that acquiescence will manifest as an overall tendency to agree with positively keyed items from orthogonal trait factors. Acquiescence indeed will affect these scales (e.g., by increasing their correlation), so that it will be confounded to the true trait. People with high observed scores because of true elevations in all measured factors will be undistinguishable from people that have their scores inflated because of acquiescence. Therefore, in those scales composed of only positively keyed items, acquiescence is confounded with content trait and cannot be properly disentangled. However, it might be identified if we have a proper number of logical antonyms measuring both ends of a construct. This view agrees with Hofstee et al. (1998) who wrote "acquiescence may be defined as the discrepancy between the average over opposites and the scale midpoint. In this definition acquiescent responding is illogical and is therefore best treated as an artifact" (p. 898).

We also stress that negatively keyed items should be constructed avoiding the word "not" ('I am not too talkative'), and rather using affirmative statements measuring the low end of a construct ('I am a bit quiet'). This will avoid the burden of higher cognitive load on negatively keyed items, when contrasted to their positively keyed counterparts.

Research on acquiescence is moving to more advanced modeling approaches using Multidimensional Item Response Theory (MIRT, see Maydeu-Olivares and Coffman 2006; Primi et al. 2019b; Savalei and Falk 2014). Nevertheless, some misconceptions still appear in the literature. Our purpose in this chapter is (a) to revise basic concepts of methods for measuring and controlling acquiescence, (b) describe some important properties of balanced scales, (c) examine if methods of controlling acquiescence yields ipsative scales (d) explain the mechanism underlying the correction of acquiescence, and (e) compare the centering and standardizing correction methods. We propose a simple simulation providing R code to illustrate the concepts with visualizations from simulation.

1.2 How to Measure and Control for Acquiescence? Re-centering Approach

Consider a six-item scale composed by three pairs of antonym items scored in a five-point Likert scale using the previously described category labels of similarity to self. Let $i = 1, 2, 3$ be positive keyed items (PK), $i = a, b, c$ be the negative ones (NK), and x_{ij} be the original response of subject j on item i. The aquiescence index acq_j of a subject j is given by:

$$\text{acq}_j = \frac{1}{6}\left[\sum_{i=1}^{3} x_{ij} + \sum_{i=a}^{c} x_{ij}\right] \tag{1}$$

Note that we are averaging items before reversing negatively phrased items to capture the overall tendency to agree with the scale categories. Since these items measure the same trait, but come from opposite ends of the trait continuum, agreement with positive items should co-occur with disagreement with negative items. Therefore, the expected score on this index will be $\text{acq}_j = 3$. If $\text{acq}_j > 3$ or $\text{acq}_j < 3$, this will indicate inconsistent responding in the form of high acquiescence or disacquiescence (i.e., disagree more than agree), respectively. Note that, in this example, all three items are logical opposites, the reason why the average over opposites is an acquiescence index acq_j.

When a subject answers "5" (completely like me) to an extraversion item as "I am often too talkative" and "3" (moderately like me) to its logical opposite "I am often too quiet," his or her acquiescence index will be $\text{acq}_j = 4$, that is, 1 point away from the scale mid-point of 3. To this difference from the scale mid-point and the acquiescence index we call discrepancy. So, in order to re-center this subject response we can add the discrepancy $3–4 = -1$ of each item response, recoding "3" into "2" and "5" into "4." In this way, recoded item means are settled back to the scale mid-point of 3. This method, proposed by Ten Berge (1999), is called re-center approach, and the recoded scores are controlled for acquiescence.

In brief, the re-centering procedure is done by: (a) calculating the acquiescence index acq_j over semantic opposite pairs for each individual j; (b) recoding original item responses by subtracting the scale's midpoint M_o of the acquiescence index: $M_o - \text{acq}_j$, and then adding this discrepancy to the original responses; (c) reversing negative items; and (d) calculating the total or average scores.

1.3 Some Properties of Balanced Scales

One interesting feature of balanced scales – those in which a positive keyed item is balanced with a negative opposite – is that their average/sum scores are automatically controlled for acquiescence. Let scr_j be the classical average/total score on the example of six items. If we do a little regrouping it will be:

$$\mathrm{scr_j} = \frac{1}{6}\left[\sum_{i=1}^{3} x_{ij} + \sum_{i=a}^{c} \left(6 - x_{jj}\right)\right] = \frac{1}{2}\left[\frac{\sum_{i=1}^{3} x_{ij}}{3} - \frac{\sum_{i=a}^{c} x_{ij}}{3}\right] + 3 \tag{2}$$

Now let scr. rec$_j$ be the individual j recoded score using the procedure outlined:

$$\begin{aligned}
\mathrm{scr.rec}_j &= \tfrac{1}{6}\left[\sum_{i=1}^{3}\left(x_{ij} + \left(3 - \mathrm{acq}_j\right)\right) + \sum_{i=a}^{c} 6 - \left(x_{ij} + \left(3 - \mathrm{acq}_j\right)\right)\right] \\
\mathrm{scr.rec}_j &= \tfrac{1}{6}\left[\sum_{i=1}^{3} x_{ij} + 9 - \sum_{i=1}^{3} \mathrm{acq}_j + 18 - \sum_{i=a}^{c} x_{ij} - 9 + \sum_{i=a}^{c} \mathrm{acq}_j\right] \\
\mathrm{scr.rec}_j &= \tfrac{1}{6}\left[\sum_{i=1}^{3} x_{ij} - \sum_{i=a}^{c} x_{ij} + 18\right] \\
\mathrm{scr.rec}_j &= \tfrac{1}{2}\left[\frac{\sum_{i=1}^{3} x_{ij}}{3} - \frac{\sum_{i=a}^{c} x_{ij}}{3}\right] + 3
\end{aligned} \tag{3}$$

Therefore, $\mathrm{scr}_j = \mathrm{scr.\ rec}_j$. In balanced scales, there is no need for additional procedures, because the classical score is automatically controlled for acquiescence.

This is a very important characteristic to remember. Total scores are controlled for acquiescence variance. However, item scores are not. When researchers run item factor analysis on scales that contain true and false keyed items, raw item score variance is a mix of true variance, acquiescence, other systematic factors, and random error. The acquiescence often distorts the factor structure, producing two factors that separate positively from negatively phrased items, even when these items are supposed to measure a unidimensional construct (see Primi et al. 2019a). Based on these results from factor analysis, some might be tempted to conclude that items cannot be summed up because they measure two different factors.

However, ironically, when summing items from balanced scales, acquiescence – that is, the core reason that distort correlations and create a two-factor structure – is partialled out, yielding a cleaned total score close to a unidimensional solution. This can be verified by running item factor analysis on item scores controlled for acquiescence using the formula in step *b* outlined above (Primi et al. 2019a; Soto and John 2017, p. 2).

1.4 Noise Canceling Mechanism

The way balanced audio cables work offers a good analogy for understanding what happens on balanced scales. Balanced audio cables use two wires to carry two copies of the audio signal from a source, for instance a microphone. But the signal polarity is reversed in one of the wires. When external noise comes along the way and interferes with the signal while it travels to the receiver, it will affect both wires/signals. Since noise interfere on both wires, this results in two copies of the noise with positive sign. The receiver device flips back the signal from the reversed wire. While reversing it will flip the voice signal from negative to positive, and it will flip noise in this wire from positive to negative. Now we end up with two copies of the noise: one negative and one positive. Finally, when the two signals are summed up in the receiver, noise cancels out, and the signal remains intact and amplified.

This is analogous to what happens with balanced personality scales: Person's true trait is the source signal we are interested in; acquiescence is the noise that comes along with the source signal. The inclusion of items that are logical bipolar opposites will make a copy of the signal with a positive sign – on true keyed items – and a negative sign – on false keyed items. Similarly, acquiescence will influence both items by introducing a positive sign (overall tendency to agree). When scoring the test, we reverse negative items and sum them up just as the audio device does with signal from two wires. Therefore, the source signal is intact and amplified, and acquiescence cancels out.

To make the analogy clearer, let us assume a simple model with no measurement errors and item effects, that is, no differences in item difficulties (see Primi et al. 2019c, for a MIRT formulation of this, conceiving acquiescence as differential person functioning). In this model, agreement with the item i by a person j, x_{ij} is a function of a person's true trait T_j and acquiescence A_j. In positively phrased items $x_{ij} = T_j + A_j$ whereas in negatively phrased items $x_{ij} = -T_j + A_j$. The core part of the formula describing the scoring procedure is the difference between positive and negative items $\sum_{i=1}^{3} x_{ij} - \sum_{i=a}^{c} x_{ij}$. So, for a balanced test with two items scr. $\text{rec}_j = 1/2\ (T_j + A_j - (-T_j + A_j)) = T_j$. It is interesting that this noise canceling mechanism was used intentionally by Mirowsky and Ross (1991) to create a measure of locus of control that have acquiescence and social desirability canceled.

1.5 Does Re-centering Produce Ipsative Scores?

Some researchers name the centering transformation as an ipsative transformation. Chan and Bentler (1998) explain what an ipsative scoring is: "Let $x = (x_1 \ldots x_p)$ be a $p \times 1$ column vectors such that $\sum_{i=1}^{p} x_i = l'x = c$ where l is a $p \times 1$ unit vector and c is constant scalar. So x is a p-dimensional data vector with ipsative property" (p. 215).

The transformation outlined above that subtracts acq_j, the subject mean, from each item score x_{ij} producing a transformed score $x'_{ij} = x_{ij} - \text{acq}_j$ is indeed an ipsative transformation because if we sum these transformed scoresx'_{ij} for all items they will sum to zero for each individual:

$$\sum_{i=1}^{3} \left(x_{ij} - \text{acq}_j\right) + \sum_{i=a}^{c} \left(x_{ij} - \text{acq}_j\right) = 0 \qquad (4)$$

Ten Berge (1999) examined properties of ipsative transformation of balanced personality scales containing NK and PK items. He noted that balanced scales are a special case with peculiar properties. The detail that needs to be remembered is that when calculating subjects' scores, we reverse half x'_{ij} item scores of NK items and then calculate average item scores. Because of this reversal, the scores are not further ipsative. There will be between-subject variance left. But this variance is disentangled from acquiescence (the variance related to acq_j). If scales are composed with only true keyed items or only false-keyed items, then we will have ipsative scores. This is the idea of Ten Berge (1999) paper's title: "A Legitimate Case of Component Analysis of Ipsative Measures, and Partialling the Mean as an Alternative to Ipsatization."

2 Simulation

2.1 What Does Correction for Acquiescence Really Do?

In order to understand the result of the acquiescence correction, we prepared a simple simulation in R.[1] We first simulated all response patterns of a balanced scale of six items: three positively keyed and three negatively keyed. All items are scored on a five-point Likert scale. This resulted in $5^6 = 15{,}625$ possible response patterns, which composed a database for the analyses that follows. Using this database, we computed scr_j, acq_j, scr. rec_j. Upper part of Figure 1 shows the scatter diagram of scr_j vs scr. rec_j colored by acq_j. It can be observed (upper part) that the classical scores are the same as controlled for acquiescence scores.

Lower part shows scr_j vs acq_j, illustrating how acquiescence correction operates. When acquiescence is equal to the expected value of 3 under a consistent responding (x-axis), that is, negative item responses reflected from positive item responses, scores have a full amplitude of variation from 1 to 5 (y-axis). As responses deviate from the expected value either because individuals are acquiescent $\text{acq}_j > 3$ or disacquiescent $\text{acq}_j < 3$, scores amplitude is shrunk. So, when individuals agree with items in an inconsistent manner, their scores are regressed toward the mid-point of the scale. In extreme cases, when individual's acquiescence is 1 (pattern 11111) or 5 (pattern 55555), scores will equal the mid-point with no variance.

Figure 1 also illustrates how acquiesce variance is being partialled out. Imagine that a sample has many acquiescent individuals with $\text{acq}_j = 4$. The amplitude of their scores will have less variance. Hence, the total variance will be less than the maximum amplitude possible for a sample of consistent responders with $\text{acq}_j = 3$. This happens because part of the item response variance is due to acquiescence and, therefore, it is partialled out.

[1] R code is available here: http://www.labape.com.br/acqu_mirt/methods_of_recoding.html see also: https://github.com/rprimi/acqu_mirt

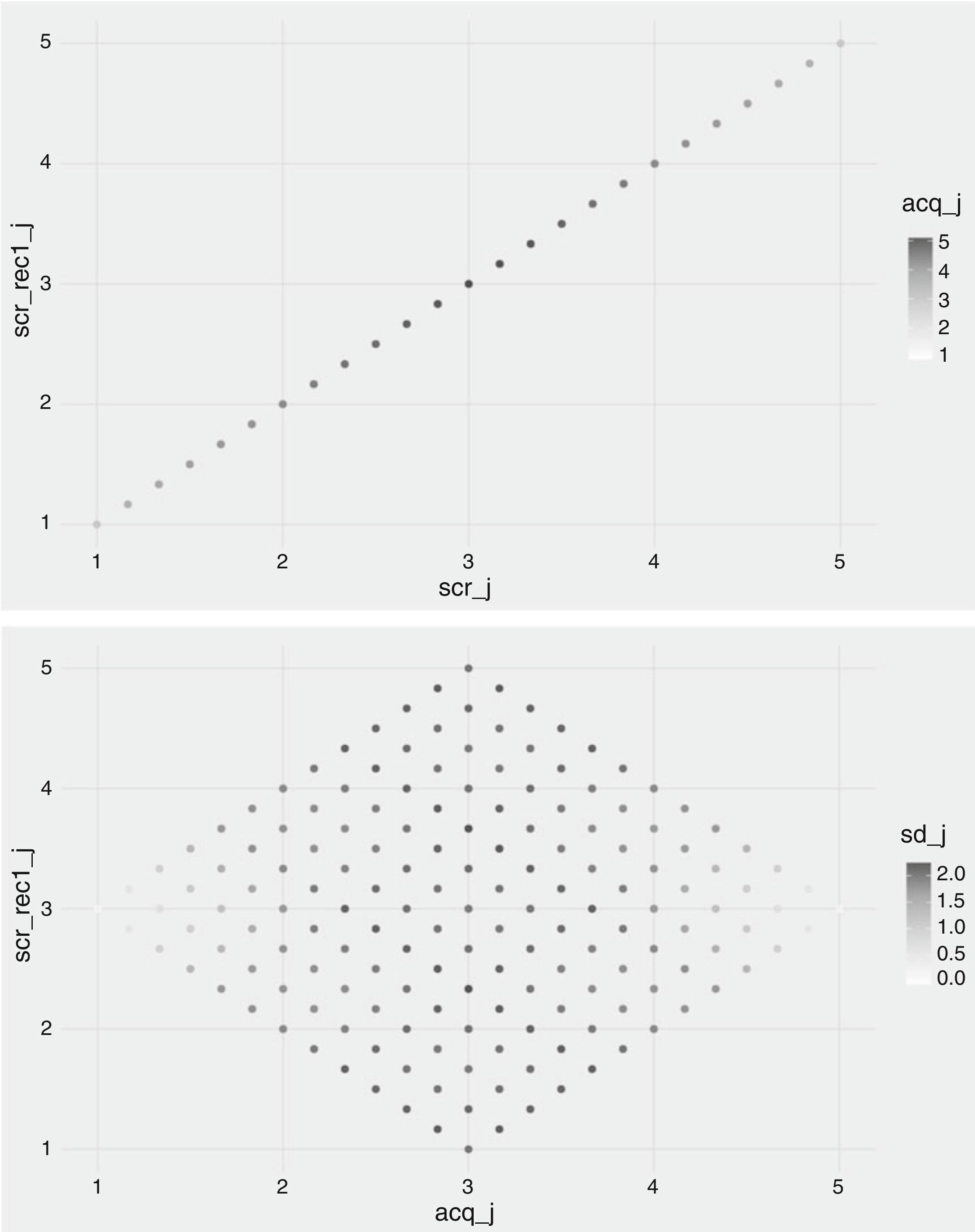

Fig. 1 Correlation of scr_j vs scr. rec_j colored by acq_j (upper panel) and scr. rec_jvs acq_j (lower panel)

2.2 *What Happens When We Center and Standardize by an Individual's Spread?*

Another method of acquiescence correction proposes that after subtracting the individual's mean on all items (acquiescence index acq_j) we divide by the individual's standard deviation. Hofstee et al. (1998) called this method "row standardization that additionally corrects for individual differences in spread," and they warn us that "contrary to the case of acquiescence a cogent rationale for this correction is lacking" (p. 901).

Using our simulated database, we computed the standardized score:

$$\mathrm{scr.z.rec}_j = \frac{1}{6}\left[\sum_{i=1}^{3}\left[\left(x_{ij} - \mathrm{acq}_j\right)/\mathrm{sd}_j\right] - 1\sum_{i=\mathrm{a}}^{\mathrm{c}}\left[\left(x_{ij} - \mathrm{acq}_j\right)/\mathrm{sd}_j\right]\right] \tag{5}$$

The formula for the individual standard deviation can be written as:

$$\mathrm{sd}_j = \sqrt{\left(\frac{\sum_1^{i=6} x_{ij}^2}{6} - \mathrm{acq}_j^2\right)} \tag{6}$$

Note that standard deviation is dependent on squared acquiescence. Figure 2 shows the relationship between the subject's standard deviation (y-axis) and his acquiescence index (x-axis) making this dependency clear. Formula (6) has a quantity plus − 1 multiplying acquiescence squared so it has the shape of an inverted parabola. Note that, as acquiescence diverge from its expected value of 3 on both directions (to 1 or 5), the standard deviation decreases.

Note that the numerator of formula (5) is the re-centering approach, which decreases the item score – i.e., makes the item score less extreme, closer to the

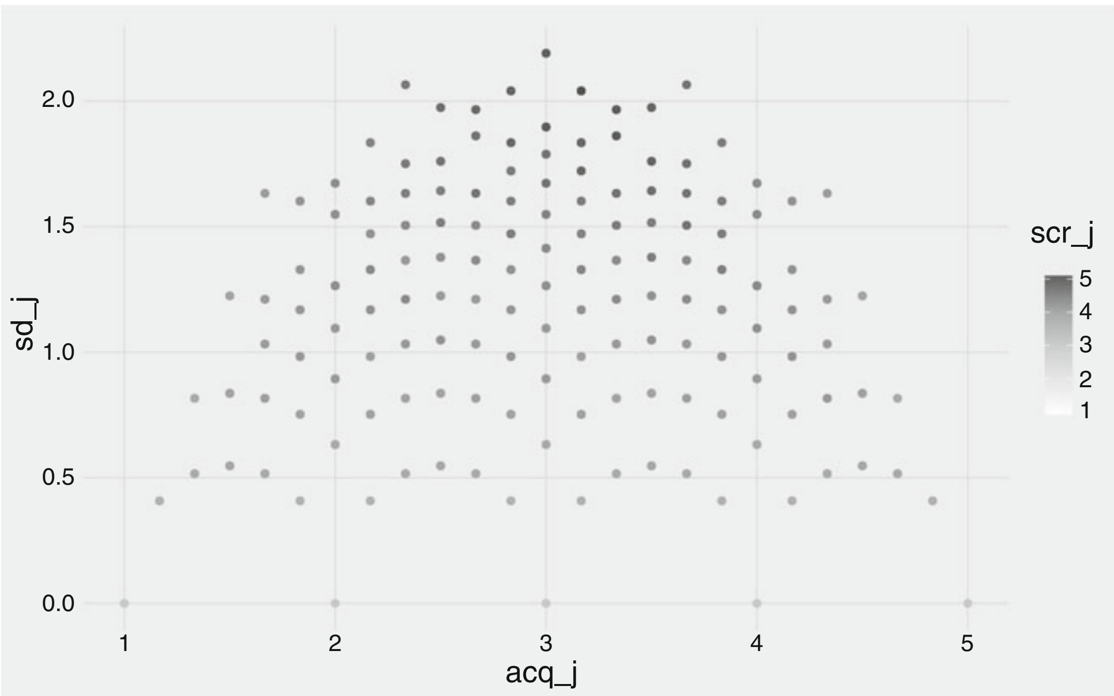

Fig. 2 Relationship between the individuals' acquiescence index (x-axis acq_j) and their standard deviation (sd_j)

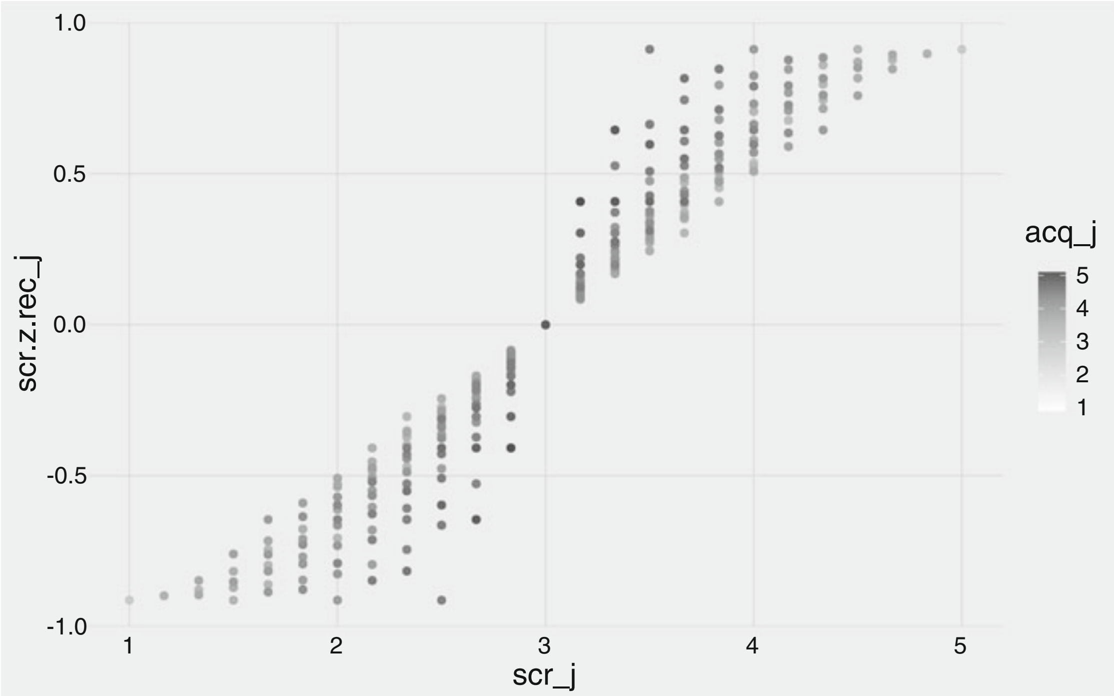

Fig. 3 Relationship between the individuals' standardized score and their original score colored by individual's standard deviation (spared of responses sd_j)

scale midpoint – proportionally to high acquiescence (or increases item score proportionally to high disacquiescence). Nevertheless, dividing this transformed score by sd_j will expand the amplitude of item score back, since high acquiescence is related to low sd_j.

Figures 3 and 4 illustrate these relationships. Figure 3 shows how the standardized score (y-axis) is related to the original score (x-axis) that is the result of re-centering approach with points colored by acquiescence. Note that individuals with high acquiescence had original scores with reduced amplitude (around 2–4) due to the automatic correction, but they are mapped onto the same amplitude -1 to 1 as individuals with expected acquiescence of 3.

Figure 4 shows standardized score on the y-axis versus acquiescence on the x-axis, similar to lower graph of Figure 1. Note that this figure does not have the diamond shape as before, meaning that the correction for acquiescence is not working properly. Even individuals with high acquiescence (or high disacquiescence) have the same amplitude on the recoded scores (y-axis). Another way to interpret this figure is that standardized scores are unrelated to acquiescence. In conclusion, standardizing misses the essence of controlling for acquiescence.

Imagine a subject A, that endorses "5" (completely like me) to an item "*I am too talkative*" and "1" (not at all like me) to "*I am too* quiet." Now imagine a subject B, that also endorses "5" (completely like me) to the first item, but "4" (a lot like me) to the second item. Subject A will have a $scr_j = 5$ and scr. z. $rec_j = 1.5$ indicating high extraversion. Subject B will have a score $scr_j = 3.5$. The model will regress the subject B score toward the mid-point due the inconsistent pattern. Is person

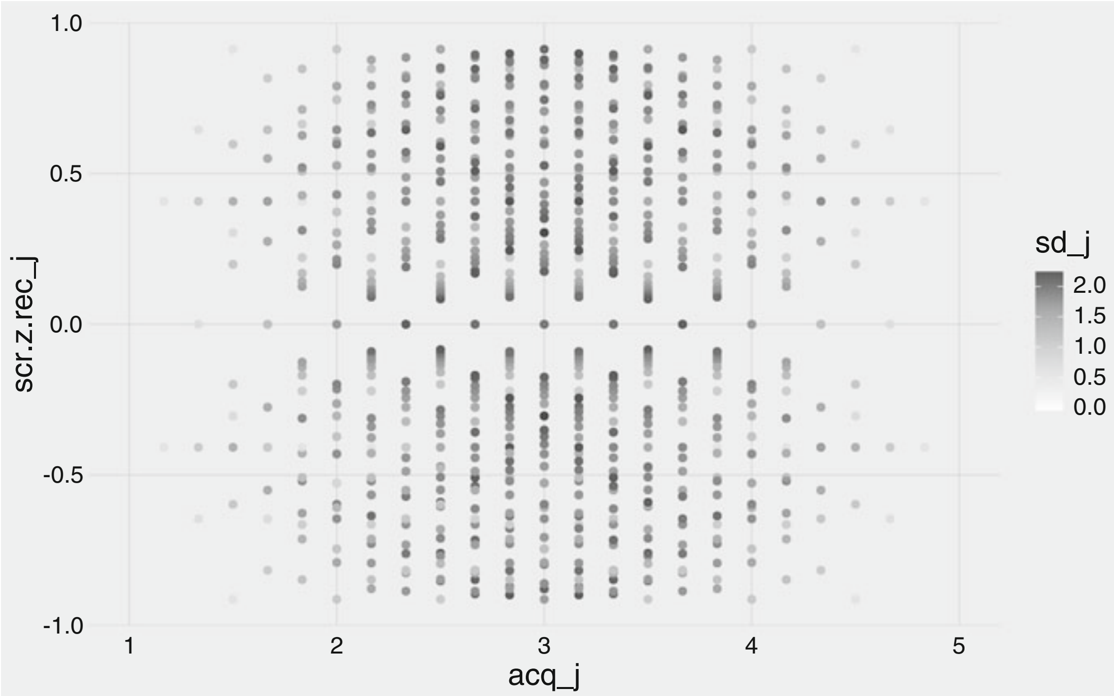

Fig. 4 Relationship between individual's standardized score and individual's acquiescence colored by the standard deviation (spared of responses sd_j)

B talkative or quiet? It is difficult to know considering his inconsistent response pattern.

Importantly, subject B standardized score will be scr. z. $rec_j = 1.5$ indicating a similar level of extraversion to subject A. The logic of standardized score is that person B uses only a tiny gradient of the full amplitude of the scale (4 and 5). Person B has a $sd_j = .5$ compared to $sd_j = 2$ of Person A. Therefore, this method considers this tiny deviation as indicative of a high level of extraversion in the context of a restricted use of the scale. This method standardizes the scale across subjects.

3 Discussion

We revised several basic properties of measuring and controlling for acquiescence. We emphasize four key conclusions that are important for research practice. First, we conclude that the mainstream unfavorable view about negatively phrased items is based on a questionable practice of internal structure analysis (Suarez-Alvarez et al. 2018). Researchers usually run item factor analysis on raw responses without controlling for acquiescence. As a consequence, results will likely show factors grouping PK items separate from NK items. This will lead to a conclusion advising that scales should be splitted into two factors or negative items should be avoided (Gehlbach and Artino 2017). Nevertheless, as we have shown, when we sum PK and NK items, a "noise canceling" mechanism operates, and part of the cause for the separation of the two factors is removed from the total score. Our conclusion advises that researchers should run item factor analysis based on recoded item scores, as this removes acquiescence systematic error, producing unbiased item-correlation matrices that, in turn, will provide better evidence about whether items conform to a unidimensional bi-polar solution (Primi et al. 2019a).

Second key point refers to the method of measuring acquiescence. We concluded that good semantic opposite pairs requiring PK vs. NK items are needed to measure aquiescence as inconsistent responding. Other operationalizations do not require NK items and propose averaging agreement over several uncorrelated dimensions of scales composed of PK items (Wetzel et al. 2016). It can be inferred from our demonstration that scales composed of items with the same key (only PK or only NK) will not create the noise canceling mechanism.

The third key point refers to the methods of correction. This study shows graphically that the correction is essentially a "partialing out" action that removes acquiescence variance from the item scores (as was explained by Ten Berge 1999). By using this procedure, we can expect the scores validity to increase at the level of individuals. However, after correction, we shrink extreme scores that regress to the mid-point because we do not have much confidence on item scores to assert individual's salience in one or another direction due to inconsistent responding. This is not exactly a more valid score. By contrast, when we study groups of individuals and correlations between measures, we do have more valid coefficients due to the clearance of a systematic error that might suppress or inflate correlations (see: Primi et al. 2019a).

Still one conclusion related to the method of correction is that centering is the method that should be used. We have shown that row standardization may undo the correction. Hofstee et al. (1998) has warned about the need of more studies for this method. We consider that studies on response process and cognitive laboratories will be important to shed light on the underlying processes of inconsistent responding or agreement behavior. If inconsistent responding over semantic opposites is due to general idiosyncratic restriction in the use of the full Likert categories, then row-standardization that equate scale use is justified. Alternatively, if inconsistent responding is more related to Messick's (1966) "interpretative acquiescence" related to verbal comprehension skills, then only the centering method should be used.

Last, we highlight the mechanism of noise canceling as a clever method for identifying and disentangling systematic error from true trait variance. Interesting examples of using this mechanism in other types of bias (like defense bias) is tested by Mirowsky and Ross (1991). This is an example of experimental manipulation of item design features to create more pure measures as it is proposed for cognitive testing by Embretson (1994).

Finally, we point to some limitations of this study. An important one is the classical test theory assumptions of equal item difficulties of antonym pairs. Another limitation is the restriction of our simulations to balanced scales only. An interesting follow-up study would be to relax these assumptions with MIRT methods and investigate the scale properties when items are unbalanced as it is done in some examples in Ferrando and Lorenzo-Seva (2010), Primi et al. (2019b) and Savalei and Falk (2014).

Acknowledgments We acknowledge the support of the Ayrton Senna Foundation. The first, second, and third authors also received a scholarship from the National Council on Scientific and Technological Development (CNPq, 310909/2017-1), Coordination for the Improvement of Higher Education Personnel (CAPES, 88881.337381/2019-01) and São Paulo Research Foundation (FAPESP, 2018/10933-8).

References

Chan, W., & Bentler, P. M. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika, 63*(4), 369–399. https://doi.org/10.1007/BF02294861.

Embretson, S. E. (1994). Applications of cognitive design systems to test development. In C. R. Reynolds (Ed.), *Cognitive assessment. A multidisciplinary perspective* (pp. 107–135). New York: Plenum Press.

Ferrando, P. J., & Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology, 63*(2), 427–448. https://doi.org/10.1348/000711009X470740.

Gehlbach, H., & Artino, A. R. (2017). The survey checklist (manifesto). *Academic Medicine, 93*(3), 1. https://doi.org/10.1097/ACM.0000000000002083.

Hofstee, W. K. B., Ten Berge, J. M. F. T., & Hendriks, A. A. J. (1998). How to score questionnaires. *Personality and Individual Differences, 25*(5), 897–909. https://doi.org/10.1016/S0191-8869(98)00086-5.

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11*(4), 344–362. https://doi.org/10.1037/1082-989X.11.4.344.

McCrae, R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19*(2), 97–112. https://doi.org/10.1177/1088868314541857.

Messick, S. (1966). The psychology of acquiescence: An interpretation of research evidence. *ETS Research Bulletin Series*, i–44. https://doi.org/10.1002/j.2333-8504.1966.tb00357.x.

Mirowsky, J., & Ross, C. E. (1991). Eliminating defense and agreement bias from measures of the sense of control: A 2 X 2 index. *Social Psychology Quarterly, 54*(2), 127–145. https://doi.org/10.2307/2786931.

Primi, R., De Fruyt, F., Santos, D., Antonoplis, S., & John, O. P. (2019a). True or false? Keying direction and acquiescence influence the validity of socio-emotional skills items in predicting high school achievement. *International Journal of Testing*. https://doi.org/10.1080/15305058.2019.1673398.

Primi, R., Hauck-Filho, N., Valentini, F., Santos, D., & Falk, C. F. (2019b). Controlling acquiescence bias with multidimensional IRT modeling. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative psychology* (pp. 39–52). Cham: Springer. https://doi.org/10.1007/978-3-030-01310-3_4.

Primi, R., Santos, D., De Fruyt, F., & John, O. P. (2019c). Comparison of classical and modern methods for measuring and correcting for acquiescence. *British Journal of Mathematical and Statistical Psychology, 72*(3), 447–465. https://doi.org/10.1111/bmsp.12168.

Savalei, V., & Falk, C. F. (2014). Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate Behavioral Research, 49*(5), 407–424. https://doi.org/10.1080/00273171.2014.931800.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://doi.org/10.1037/pspp0000096.

Suarez-Alvarez, J., Pedrosa, I., Lozano, L. M. B., Garcıa-Cueto, E., Cuesta, M., & Muñiz, J. G. F. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33.

Ten Berge, J. M. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research, 34*(1), 89–102. https://doi.org/10.1207/s15327906mbr3401_4.

Wetzel, E., Ludtke, O., Zettler, I., & Bohnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment, 23*, 279–291. https://doi.org/10.1177/1073191115583714.