RESEARCH   Open Access

Construct validity and reliability of Olweus Bully/Victim Questionnaire – Brazilian version

Francine Guimarães Gonçalves[*], Elizeth Heldt, Bianca Nascimento Peixoto, Gabriela Adamatti Rodrigues, Marcelly Filipetto and Luciano Santos Pinto Guimarães

Abstract

The Revised Olweus Bully/Victim Questionnaire (OBVQ) is among the few bullying assessment instruments with well-established psychometric properties in different countries. Nevertheless, the psychometric properties of the Brazilian version (Questionário de Bullying de Olweus - QBO) have not been determined. We aimed at verifying the construct validity and reliability of the bully and victim scales of the QBO. To achieve that goal, the victim and bully scales were assessed using polytomous item response theory (IRT). The best fit was obtained with a generalized partial credit model that is capable of measuring the specific discriminating power for each item in these scales. The QBO was administered to 703 public school students (mean age: 13 years; standard deviation = 1.58). Based on IRT analysis, the number of response categories in each item was reduced from four to three. Cronbach reliability scores were satisfactory: $\alpha = 0.85$ (victim scale) and $\alpha = 0.87$ (bully scale). In this study, hurtful comments, persecution, or threats had high power to discriminate victims and bullies. For both QBO scales, higher severity parameters were observed for direct bullying items. The results also show that the construct of both QBO scales measures the same construct proposed for the overall instrument. Thus, the QBO can be administered to different Brazilian populations to assess the main characteristics of bullying: repetition of behavior over time and intentionally acting to humiliate, threaten, or harm somebody.

Keywords: Bullying, Psychometrics, Construct, Reliability

Background

Bullying, one of the most common forms of violence in schools, is defined as power asymmetry associated with differences in age, gender, or race which is exploited by one or more individuals with the intention of hurting or humiliating another (Olweus, 1993). Recurrence over time is also a key aspect of bullying (Berger, 2007), along with the involvement of a bully, or perpetrator, and of a victim, the target of the aggression. Some individuals may be at the same time perpetrators and victims, and are therefore classified as bully-victims (Malta et al., 2010).

In broad terms, bullying may be classified as direct or indirect (Lopes Neto, 2005). Direct, face-to-face bullying draws more attention because it involves open aggression, including public verbal abuse, intentional exclusion

from groups, punching or pushing, or other types of physical aggression. Indirect bullying involves spreading negative rumors or accusations about a person who is not present to defend him or herself, or indirect negative comments in the presence of the target (Lopes Neto, 2005).

There is often only a thin line between "normal" and healthy teasing between peers and behaviors that tend to be classified as bullying (Volk et al. 2012). In fact, bullying is understood as a social phenomenon rather than a psychiatric disorder (Lopes Neto, 2005). Nevertheless, studies have shown that bullying has a severe negative impact on academic performance (Webster-Stratton et al. 2008), with consequences that may extend into adulthood for both victims and perpetrators (Malta, et al., 2010).

Bullying is usually assessed based on self-report instruments (Kert et al. 2010). Therefore, the findings of a systematic review of 31 articles describing 27 self-report instruments used to evaluate bullying are a reason for

* Correspondence: fran_alvess@hotmail.com
Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Gonçalves *et al. Psicologia: Reflexão e Crítica* (2016) 29:27

Page 2 of 8

concern (Vessey et al. 2014) – the review reports only "limited evidence supporting the reliability, validity, and responsiveness of existing youth bullying measures" (pg. 819). That finding challenges the usefulness of self-report to assess bullying (Vessey, et al., 2014). In this context, determining the validity and reliability of these instruments, that is, the extent to which they discriminate bullying from normative peer conflicts, is crucial to ensure that data reflect the trends of phenomena under observation. This is also true for translation and cultural adaptations of instruments to different languages.

Among the instruments cited in the systematic review by Vessey et al. (2014), the Revised Olweus Bully/Victim Questionnaire (OBVQ) is among the few with well-established psychometric properties in different countries (Kyriakides et al. 2006). The OBVQ contains two separate scales, one focusing on acts of victimization and one focusing on acts of bullying. The answers to each question are chosen from a multiple choice Likert scale, an aspect that has been criticized. According to Kyriakides et al. (2006), the use of a Likert scale "disregards the subjective nature of the data by making unwarranted assumptions" about the meaning of each choice because "the relative value of each response category across all items is treated as being the same" (p. 784). To circumvent this limitation, the authors propose the use of Item Response Theory (IRT), a mathematical method that analyzes the scores in relation to each other in order to evaluate whether the instrument is indeed capable of achieving its goal in a universal manner, that is, across different populations. Kyriakides et al. (2006) found that the OBVQ had satisfactory psychometric properties in a Greek sample (construct validity and reliability). That study also encourages the use of IRT to test other cultural adaptations of the OBVQ.

A Brazilian Portuguese version of the OBVQ, *Questionário de Bullying de Olweus* (QBO), is also available. The QBO contains 23 items that investigate the frequency with which individuals experience and/or engage in bullying behaviors 30 days before the survey (Olweus, 1996; Fischer et al., 2010). Subjects who experience or perpetrate any of the behaviors at least three times a month are classified as victims or bullies respectively. However, the psychometric properties of the QBO have not yet been determined.

In light of the above, the aim of the present study was to verify the construct validity and reliability of the bully and victim scales of the QBO using an IRT model.

Method

This methodological study was approved by the Research Ethics Committee of the Federal University of Rio Grande do Sul, and by the Municipal Department of Health of the city of Porto Alegre (protocol number: CAAE

19651113.5.0000.5338). All parents and/or guardians provided written consent for the participation of their children in the study, and all adolescents signed an assent form prior to enrollment.

Participants

Fifth to ninth-grade students of both sexes, aged between 10 and 17 years, attending three public schools from the city of Porto Alegre (RS, Brazil), were eligible for enrollment. A total of 713 agreed to participate and were recruited. Of these, 10 were excluded based on teacher report of intellectual disability. Thus, the final sample included 703 (98.6 %) adolescents, of whom 380 (54 %) were girls. Mean age was 13 years (± standard deviation, SD, 1.58 years). Race was self-reported as white ($n = 308$; 43.8 %), brown ($n = 194$; 27.6 %), or black ($n = 173$; 24.3 %).

Instrument

The questionnaires were administered during school hours, in the presence of two members of the research team who had been previously trained in the use of these instruments.

The QBO is a self-report instrument composed of 23 items about bullying (bully scale) and 23 items about victimization (victim scale). Each item describes a different behavior, and the respondent is asked to determine the frequency with which this behavior occurred over the past month. For instance: "Dei socos, pontapés ou empurrões/I hit, kicked or pushed someone" (bully scale); "Me deram socos, pontapés ou empurrões/I was hit, kicked or pushed" (victim scale).

Participants choose a response to each of the 23 items from a four-category Likert scale that reflects the frequency of behaviors: (1) "Nunca/Never", (2) "Uma ou duas vezes no mês/Once or twice a month", (3) "Cerca de uma vez por semana/Around once a week", and (4) "Várias vezes por semana/Several times a week" (Olweus, 1996; Fischer, et al., 2010). Because the QBO employs multiple-choice answers, it is said to be polytomously scored.

Data analysis procedures

Polytomous item response theory (IRT) was used to determine QBO validity. A discriminating parameter is calculated for each item. This parameter reflects the influence of each item on the latent variable – the higher the discriminating parameter, the higher the relevance of the item for the proposed measurement. A severity parameter is also determined, reflecting to which degree the behavior is enacted – a student with a high severity parameter is more likely to choose the highest score for the item in the Likert scale (Andrade et al. 2000). Two IRT models were tested to determine which was best suited

Gonçalves *et al. Psicologia: Reflexão e Crítica* (2016) 29:27

Page 3 of 8

to assess the ability of each item to identify victims and bullies: graded response (GRM) and generalized partial credit models (GPCM), described by Samejima (1969) and Muraki (1992) respectively.

Graded response model

The GRM was developed by Samejima (1969) and deals with polytomous categories arranged in ascending order. This model estimates the probability that an individual will select a given response or a higher-level one in each item on the scale based on the formula

$$P_{1,k}^{+}(\theta_j) = \frac{1}{1 + e^{-1,702*a_i(\theta_j - b_{i,k})}}$$

Where: i represents a given item in the questionnaire; j refers to the subject under assessment; k designates an item response category; n is the number of subjects in the sample; m_i is the number of response categories i; a_i is the discriminating parameter of item i, and $b_{i,k}$ is the severity parameter of response category k in item i (Andrade et al., 2000).

Generalized partial credit model

The GPCM, developed by Muraki (1992), is used to determine the discriminating power for each set of adjacent choices in a Likert scale. The GPCM is said to be "generalized" because it does not assume that all items have uniform discriminating power. The formula used for the GPCM is

$$P_{i,k}(\theta_j) = \frac{\exp\left[\sum_{u=0}^{k} D_{a_i}(\theta_j - b_{i,u})\right]}{\sum_{u=0}^{m_i} \exp\left[\sum_{v=0}^{u} 1,702 * a_i(\theta_j - b_{i,v})\right]}$$

Where: i represents a given item in the questionnaire; j refers to the subject under assessment; k designates an item response category; n is the number of subjects in the sample; m_i is the number of response categories i; a_i is the discriminating parameter of item i; and $b_{i,k}$ is the severity parameter of response category k for item i (Andrade et al., 2000).

Statistical analysis

The construct validity of the QBO was established through IRT analysis using the GRM and GPCM, both of which deal with polytomous variables. The bully and victim scales of the questionnaire were independently analyzed.

The GPCM was run in three variations: a) constant discriminating power equal to 1; b) constant discriminating power not equal to 1; and c) variable discriminating power across all items. The GRM was run in two variations: d) constant discriminating power across items;

and e) variable discriminating power across items (Andrade et al., 2000).

The best model was selected based on the comparison of the area under the curve generated by each model, where the size of the area reflects the amount of information included in the calculations. The curves with the largest area correspond to the best models. The intersections of item characteristic curves (ICC) were then analyzed to verify whether any categories should be removed from the model. Any categories with response probabilities below those of the other categories were excluded.

To facilitate the interpretation of victim and bully scores, which are normally distributed with a mean of zero and a standard deviation of one, scores were multiplied by the standard deviation of total scores, and added to the mean total score on the scale (Pasquali, & Primi, 2003). The unidimensionality of the scales (that is, the ability of scale items to measure the aspect they propose to measure, being a victim or a bully) was verified through factorial and parallel analysis. The reliability of each scale was estimated using Cronbach's alpha. A Cronbach coefficient > 0.70 indicates a satisfactory level of reliability (Pilatti et al. 2010).

Data analysis was performed using the *R* statistical software package (Gentleman, & Ihaka, 2015) and IRT analysis was performed using the *Itm* package (Rizopoulos, 2006).

Results

Firstly, the QBO with its four response categories was analyzed using the proposed IRT models, and the resulting performance curves were compared (Table 1). After that, analysis of the ICC led to the combination of response categories "Cerca de uma vez por semana/Around once a week" and "Várias vezes por semana/Several times a week." That was true for all 23 items. Figure 1 shows the graphs corresponding to item 1 in both scales, modeled with four (Fig. 1a) and three alternatives (Fig. 1b). Because the curves corresponding to items 2 to 23 were very similar to the graph plotted for item 1, they are not shown.

Because the probability that an adolescent would select category 3 ("Cerca de uma vez por semana/Around once a week") was zero in both scales of the QBO, both the victim and the bully scales were reformulated to include only three response categories: (1) "Nunca/Never", (2) "Uma ou duas vezes por mês/Once or twice a month", (3) "Uma ou mais vezes por semana/Once or more than once a week". All 23 items were recoded accordingly.

After this change, participant scores were reanalyzed using the five mathematical models. The results of this procedure are shown in Table 1. For the bully scale, the largest area was enclosed by the GPCM with non-uniform discriminative power (area = 97.8 %). In the

Gonçalves *et al. Psicologia: Reflexão e Crítica* (2016) 29:27

Page 4 of 8

Table 1 Performance curves for five mathematical models assessing the use of three and four response categories in the Brazilian Portuguese version of the Olweus Bully/Victim Questionnaire (QBO)

Models[a]	Victim scale		Bully scale	
	Four categories	Three categories	Four categories	Three categories
Generalized Partial Credit Model (GPCM)				
a) Discriminating power = 1	94.9	87.1	87.6	77.1
b) Constant discrimination	98.0	94.8	98.2	97.3
c) Variable discrimination	98.6	93.4	98.5	97.8
Graded Response Model (GRM)				
d) Constant discrimination	88.0	91.6	85.0	92.2
e) Variable discrimination	86.4	89.7	91.5	94.9

[a]Values represent the area under the curve in %

victim version, the largest area, by a small margin, was that enclosed by the GPCM with constant discriminating power (area = 93.4 %). The GPCM with variable discriminating power was selected as the best model for both versions of the scale for two reasons: firstly, the difference between its area and that of the GPCM with constant discriminating power was very small.

Secondly, this model had been selected as the most adequate for this data set since the beginning of the analysis (area = 94.8 %).

As previously described, participant scores were also transformed into standard deviation units (SD). The scores of each participant were added up to a total value, whose mean and standard deviation were calculated for
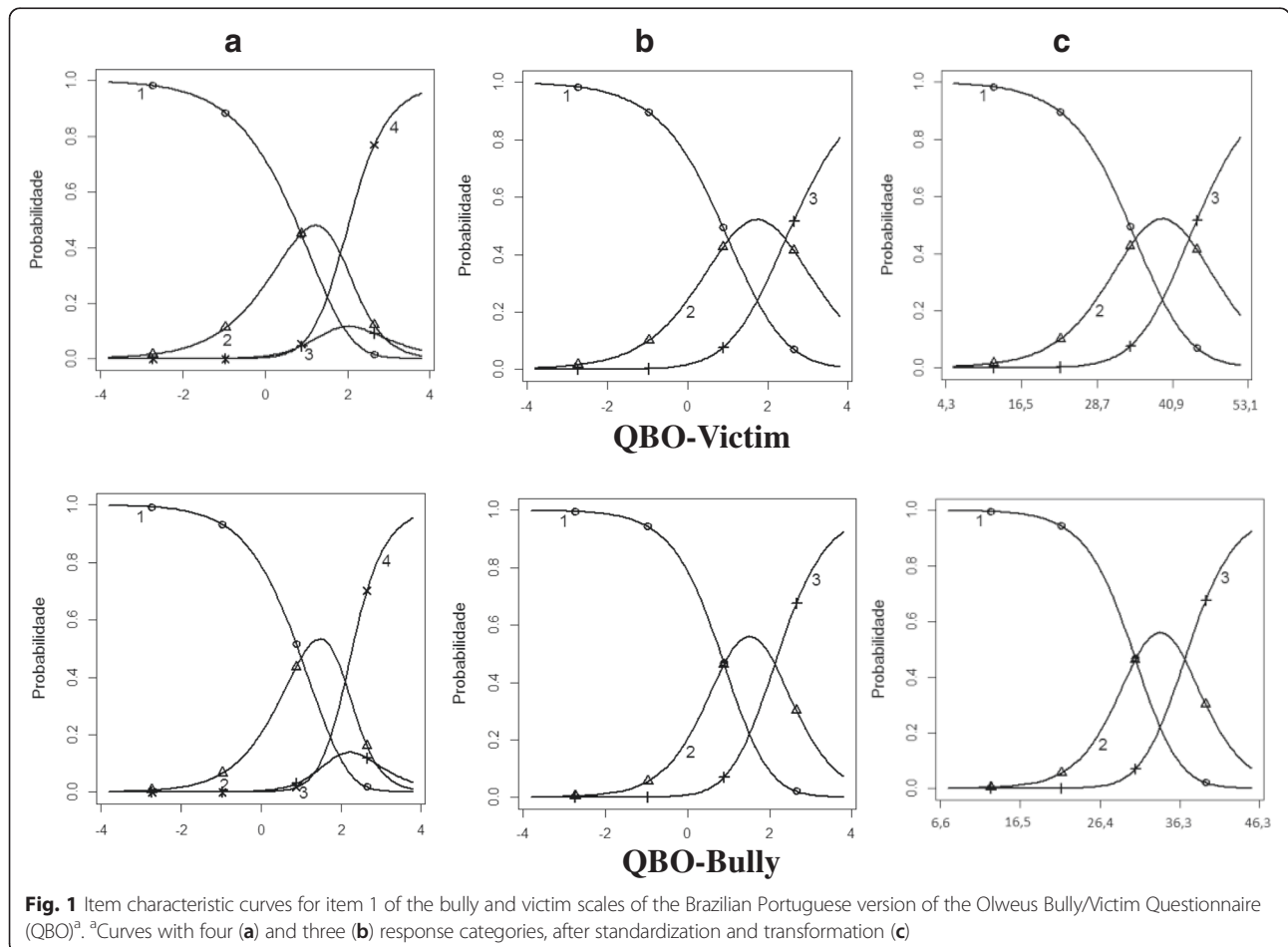
Fig. 1 Item characteristic curves for item 1 of the bully and victim scales of the Brazilian Portuguese version of the Olweus Bully/Victim Questionnaire (QBO)[a]. [a]Curves with four (**a**) and three (**b**) response categories, after standardization and transformation (**c**)

Gonçalves *et al. Psicologia: Reflexão e Crítica* (2016) 29:27

Page 5 of 8

the sample. All items from both versions of the scale were transformed using the following formula (Revelle, 2015):

$$\text{Victim version}: (\text{Raw score} \times \text{SD}) + \text{mean}$$
$$= (\text{Raw score} \times 6.1) + 28.7 \quad (1)$$

$$\text{Bully version}: (\text{Raw score} \times \text{SD}) + \text{mean}$$
$$= (\text{Raw score} \times 5.0) + 26.4 \quad (2)$$

The response parameters for item 1 following transformation are shown in Fig. 1c. The topmost curve indicates the most likely response by participants in different score intervals. As can be seen in Figure 1, in the first item of the victim version of the questionnaire, adolescents with total scores up to 34.9 were likely to respond "Nunca/Never" (Category 1); those with scores between 34.9 and 43.6 were likely to respond with "Uma ou duas vezes no mês/Once or twice per month." Finally, subjects with scores above 43.6 were most likely respond with "Uma ou mais vezes por semana/Once or more than once

a week". The same results were observed in the first item of the bully scale.

Once values were transformed to facilitate their interpretation, item parameters were evaluated. Table 2 shows the discriminating power and severity parameter for each item. The higher the discriminating power, the greater the contribution of the item to classifying the respondent as a victim or bully. In the victim version of the questionnaire, items 20, 15, and 3 were most discriminative, while items were 11, 4, 5, and 8 had the lowest discriminating power. The intersection between response categories 1 and 2 in items 4, 5, 8, 10, 11, 14, 16, and 22 suggests that the number of response categories for these items could be further reduced to two.

The most discriminative items in the bully scale (Table 3) were items 22, 15 and 3, while the least discriminative items were 23 and 6. In items 4, 5, 10, 14, 16 and 23 the high degree of intersection between categories 1 and 2 also suggested that the number of possible response categories could have been reduced to two.

Table 2 Generalized Partial Credit Model (GPCM) of items in the victim scale of Brazilian Portuguese version of the Olweus Bully/Victim Questionnaire (QBO), assuming variable discriminating power across items

Item		Severity of response category		Discrimination	Reliability[c]
		1 to 2[a]	2 to 3[a]		
20	Somebody said bad things about me or my family	32.5	38.7	2.13	0.84
15	I was followed inside or outside the school	40.1	44.7	1.88	0.85
3	I was threatened	34.1	40.5	1.77	0.85
12	People laughed and pointed at me	34.3	39.8	1.57	0.84
19	Somebody falsely accused me of snitching things from my classmates	35.8	40.6	1.48	0.84
21	Somebody tried to make other dislike me	34.8	40.4	1.47	0.85
7	Somebody yelled at me	31	38.6	1.43	0.84
23	Somebody used the Internet or a cell phone to harm/offend me	41.3	43.3	1.42	0.85
13	Somebody gave me nicknames I didn't like	32.1	36.9	1.33	0.84
14	I was cornered/pushed against a wall[b]	45.8	41.7	1.28	0.85
9	I was insulted because of a physical characteristic	35.8	38.2	1.23	0.85
17	I was not allowed to join a group of classmates	38.9	43.5	1.19	0.85
18	I was totally ignored by others	37.8	45.1	1.19	0.85
2	Somebody pulled my hair or scratched me	38.9	44.9	1.11	0.85
1	Somebody punched, kicked, or pushed me	34.9	43.6	1.1	0.85
10	I was humiliated because of my sexual preference of mannerisms[b]	50	39.9	1.09	0.85
16	I was sexually harassed[b]	51.1	41.3	1.04	0.85
22	I was forced to physically harm a classmate[b]	48.7	41.8	1.03	0.85
6	Somebody broke my things	36.4	43.6	1.01	0.86
8	I was insulted because of my color or race[b]	43.6	41.9	0.96	0.85
5	Somebody snatched my money or belongings without my consent[b]	46.9	45.5	0.92	0.85
4	I was forced to hand over my money or belongings[b]	55.2	43.1	0.88	0.85
11	Somebody made fun of my accent[b]	47.7	41.9	0.86	0.85

[a]Values correspond to intersections between response categories
[b]Items with a high degree of intersection between categories 1 and 2
[c]Cronbach's α coefficient excluding each item

Table 3 Generalized Partial Credit Model (GPCM) of items in the bully scale of the Brazilian Portuguese version of the Olweus Bully/ Victim Questionnaire (QBO), assuming variable discriminating power across items

Item		Severity of response category		Discrimination	Reliability[c]
		1 to 2[a]	2 to 3[a]		
22	I forced someone to hit/offend another classmate	36.4	37.5	3.1	0.86
15	I followed someone inside or outside the school	36.2	37.1	2.45	0.86
3	I threatened someone	33.3	37	2.43	0.85
4	I forced somebody to give me their money or belongings[b]	38	36.4	2.34	0.86
10	I humiliated somebody because of their sexual preference or mannerism[b]	37.5	37.3	2.24	0.86
13	I made nicknames for others that they didn't like	31.2	35.7	2.24	0.86
8	I insulted someone because of their skin color or race	36.4	38.4	2.2	0.86
9	I insulted someone because of a physical characteristic	34.2	36.6	2.18	0.86
14	I cornered or pushed someone against a wall[b]	36.6	36.4	2.11	0.86
20	I said bad things about someone or their family	35.2	36.7	2.08	0.86
19	I falsely accused someone of taking the belongings of classmates	36	39.4	1.97	0.88
12	I laughed or pointed at someone	31.9	36.3	1.94	0.86
21	I tried to make people dislike someone	36.2	38.5	1.84	0.86
16	I sexually harassed someone[b]	39.2	37.1	1.76	0.86
11	I made fun of someone because of their accent	35.8	37	1.66	0.86
2	I pulled someone's hair or scratched them	34.1	38.3	1.58	0.86
17	I didn't let someone join a group of classmates	34.9	37.1	1.57	0.86
1	I hit, kicked, or pushed someone	30.8	36.9	1.52	0.86
5	I snitched money or things from others[b]	41.4	36.4	1.32	0.86
7	I yelled at someone	28.1	34.3	1.32	0.86
18	I completely ignored someone	34.2	37.6	1.12	0.86
23	I used the Internet or cell phone to harm/offend a classmate[b]	40.5	37.7	0.96	0.86
6	I damaged other people's belongings	38	38.1	0.82	0.86

[a]Values correspond to the intersection between response categories
[b]Items with a high degree of intersection between categories 1 and 2
[c]Cronbach's α coefficient excluding each item

Once final scores were developed for the three-category version of the scale, using the aforementioned discriminating and severity parameters, the mean (standard deviation) of victim scores was 29.3 (SD = 5.39). The reliability (Cronbach alpha) of this scale was α = 0.85. The bully scale had a mean score of 26.8 (SD = 3.92) and a reliability of α = 0.87. The reliability of each item is shown in Table 2 (victim scale) and Table 3 (bully scale).

Unidimensionality analysis revealed that the first factor of the victim QBO scale explained 26.27 % of the variance, whereas the first factor of the bully QBO scale explained 31.05 % of the variance. A full Brazilian Portuguese version of the validated QBO appears in Additional files 1 and 2.

Discussion
The aim of the present study was to determine construct validity (using IRT) and reliability of the QBO. The findings showed satisfactory validity and reliability for both bully and victim scales of the QBO.

Given the complexity associated with the assessment of bullying, and the lack of validated instruments to evaluate this construct, the use of IRT to investigate the construct validity of both scales of the QBO, define the adequate number of response categories, and verify item discriminating power and severity was an important contribution to the literature. A recent review of 25 Brazilian articles found that in most studies involving the assessment of bullying, this phenomenon is identified using measures developed by the researchers themselves or with unknown validity for the Brazilian populations. The authors concluded that the absence of validated instruments for this purpose is a significant methodological limitation (Alckmin-Carvalho, et al., 2014). The use of IRT to determine construct validity is useful to assess latent traits, such as anxiety level, stress, and quality of life, which correlate with different items in an assessment measure. A relationship is expected between the presence of a particular condition and certain latent traits (Andrade, et al., 2000; Sartes & Souza-Formigoni, 2013).

Gonçalves *et al. Psicologia: Reflexão e Crítica* (2016) 29:27

Page 7 of 8

The present results revealed the need to combine response categories 3 and 4, so that only three response categories were kept in both scales (victim/bully) of the QBO: (1) "Nunca/Never", (2) "Uma ou duas vezes no mês/Once or twice a month", (3) "Uma ou mais vezes por semana/ Once or more than once a week". Although some items could be further modified to include only two response categories, the three alternatives were maintained for all items to ensure uniformity between the bully and victim scales. The presence of multiple categories allows for an estimation of behavior frequency, which is especially important since repetition is a core feature of bullying (Malta et al., 2010). Thus, use of the IRT model confirmed that the behaviors measured by the scale are expressions of the underlying construct, and also allowed us to determine the performance of each item of the QBO construct for Brazilian adolescents (Andrade et al., 2000; Pasquali, & Primi, 2003). As previously mentioned, a Greek study employed a similar model to evaluate construct validity and reliability of a cultural adaptation of OBVQ. That study also found satisfactory psychometric properties for both victim and bully questionnaires (Kyriakides et al. 2006).

Our findings also revealed that the items in the QBO differ in their loading to the latent variables in question. In this population, being the object of hurtful comments, persecution, or threats had high power to discriminate victims of bullying. Conversely, forcing people to be physically aggressive to others, persecuting students inside or outside the school, and issuing threats were most likely to identify bullies. These results are in line with the defining feature of bullying, which is the intention to humiliate, threaten, and harm (Olweus, 1996, Berger, 2007).

The items with the least discriminant ability for bullying victims were: being teased and being forced to hand over money or belongings, or having those taken without consent, and being humiliated in association with skin color or ethnicity. The least discriminating items in the bully scale were damaging the belongings of others and using the Internet to hurt others (cyberbullying). The fact that being teased figures among the least discriminative items for bullying victims suggests that this type of behavior may be interpreted as a friendly exchange between peers rather than an attempt to cause harm or humiliate (Volk et al., 2012).

Discriminating power is used to indicate that item estimates will remain relatively constant in future applications (Sartes & Souza-Formigoni, 2013). Concerning the QBO, that means that items with more strength to discriminate victims or bullies in our culture would be useful to assess bullying in schools in other samples of Brazilian adolescents.

The severity parameter is related to another central characteristic of bullying – the frequency of behaviors (the higher the severity parameter, the more frequent the behavior). For both, bullies and victims, the highest severity parameters were observed for direct bullying items; for bullies, the highest severity parameters were recorded for "I snitched money or things from others," "I used the Internet or cell phone to harm/offend a classmate," and "I sexually harassed someone". For victims, the highest severity parameters were observed for "I was forced to hand over my money or belongings," "I was sexually harassed," "I was forced to physically harm a classmate," and "I was humiliated because of my sexual preference of mannerisms". Also, the results show satisfactory reliability of the final scores, with α > 0.85 for both the victim and bully QBO scales.

The present study had some limitations. Although the replication of our method by other researchers is extremely desirable, we were unable to develop a syntax of our procedures for use in other statistical packages. Additionally, we did not provide a cutoff for the classification of bullies or victims. Nevertheless, the scores obtained by other samples on the victim and bully scales of the QBO can be calculated using IRT parameters estimated from our original data through the interactive method and tutorial available on the website www.professor.ufrgs.br/eheldt, in files *model_vit.Rdata* and *model_agr.Rdata*.

We found that simply adding up the scores on all items of the QBO without considering the relative weight of each item may interfere with the validity of this measure and, consequently, with the findings of studies which use the traditional versions of the QBO. Given the relevance of this topic, it is important that future studies continue to investigate the psychometric properties of this instrument, using factor analysis, for instance, to verify whether additional dimensions of bullying (e.g. direct and indirect bullying) can be identified using the QBO. Future studies focusing on the development of effective tools to identify and define the types of bullying behavior present in different samples will be essential to guide the implementation of prevention programs targeting bullying in school environments.

Conclusions

We found that simply adding up the scores on all items of the OBVQ without considering the relative weight of each item may interfere with the validity of this measure and, consequently, with the findings of studies which use the traditional versions of the OBVQ. Given the relevance of this topic, it is important that future studies continue to investigate the psychometric properties of this instrument, using factor analysis, for instance, to verify whether additional dimensions of bullying (e.g. direct and indirect) can be identified using the OBVQ.

Future studies which develop effective tools to identify and define the types of bullying behavior present in a given sample will be essential to allow for the implementation of prevention programs targeting bullying in school environments.

Gonçalves *et al. Psicologia: Reflexão e Crítica* (2016) 29:27

Page 8 of 8

Additional files

Additional file 1: Questionário de Bullying de Olweus – Vítima. (DOCX 37 kb)

Additional file 2: Questionário de Bullying de Olweus – Agressor. (DOCX 38 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FGG and EH involved in study design, data collection, literature review and manuscript drafting. BNP and MF involved in study design, data collection, data entry, and literature review for the introduction section. GR involved in study design, data collection, data entry, and literature review for the discussion section. LG responsible for statistical analysis, contributed to the methods and results sections. All authors approved the final version of the manuscript.

References

Alckmin-Carvalho F, Izbicki S, Fernandes LFB, Melo MHS. Estratégias e instrumentos para a identificação de *bullying* em estudos nacionais. Avaliação Psicol. 2014;13(3):343–50.

Andrade DF, Tavares HR, Valle RC. Teoria da Resposta ao Item: conceito e aplicações. In: XIV Simpósio Nacional de Probabilidade e Estatística. São Paulo: Associação Brasileira de Estatística; 2000. http://www.ufpa.br/heliton/arquivos/LivroTRI.pdf. Retrieved in 22 Nov 2014.

Berger KS. Update on bullying at school: Science forgotten? Dev Rev. 2007;27(1):90–126. doi:10.1016/j.dr.2006.08.002.

Fischer RM, Lorenzi GW, Pedreira LS, Bose M, Fante C, Berthoud C, Moraes EA, Puça F, Pancinha J, Costa MRRC, Vieira PF, Oliveira CPU. Relatório de pesquisa: *bullying* escolar no Brasil. Centro de Empreendedorismo Social e Administração em Terceiro Setor (Ceats) e Fundação Instituto de Administração (FIA). 2010. https://www.ucb.br/sites/100/127/documentos/biblioteca1.pdf. Retrieved in 22 Nov 2014.

Gentleman R, Ihaka R. The R Project for Statistical Computing. 2015. http://www.r-project.org. Retried in 22 Jul 2015.

Kert A, Codding R, Tryon G. Impact of the word "bully" on the reported rate of bullying behavior. Psychol Sch. 2010;47(2):193–204. doi:10.1002/pits.20464.

Kyriakides L, Kaloyirou C, Lindsay G. An analysis of the Revised Olweus Bully/Victim Questionnaire using the Rasch measurement model. Br J Educ Psychol. 2006;76:781–801. doi:10.1348/000709905X53499.

Lopes Neto AA. Bullying – aggressive behavior among students. J Pediatr (Rio J). 2005;81(5):164–72. doi:10.1590/S0021-75572005000700006.

Malta DC, Silva MAI, Mello FCM, Monteiro RA, Sardinha LMV, Crespo C, Carvalho MGO, Silva MMA, Porto DL. Bullying in Brazilian schools: results from the National School-based Health Survey (PeNSE), 2009. Cien Saude Colet. 2010;15(2):3065–76.

Muraki EA. Generalized partial credit model: application of an EM algorithm. Appl Psychol Meas. 1992;16(2):159–76. doi:10.1177/014662169201600206.

Olweus D. Bullying at school. What we know and what we can do. Oxford UK and Cambridge USA: Blackwell; 1993.

Olweus D. The Revised Olweus Bully/Victim Questionnaire. Bergen: Research Center for Health Promotion; 1996.

Pasquali L, Primi R. Basic theory of Item Response Theory (IRT). Avaliação Psicol. 2003;2(2):99–110.

Pilatti LA, Pedroso B, Gutierres GL. Psychometrics properties of measurement instruments: a necessary debate. Rev Bras Ensino Ciênc Tecnol. 2010;2(1):81–91.

Revelle W. Procedures for psychological, psychometric, and personality research. 2015. http://personality-project.orgwww.personality-project.org/r/psych/psych-manual.pdf. Retried in 15 Jul 2015.

Rizopoulos D. ltm: An R package for latent variable modeling and item response theory analyses. J Stat Softw. 2006;17(5):1–25.

Samejima F. Estimation of latent ability using a response pattern of graded scores. (Psychometric Monograph No. 17). Richmond: Psychometric Society; 1969. https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf. Retrieved in 22 Nov 2014.

Sartes LMA, Souza-Formigoni MLO. Avanços na Psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. Psicol Reflexão Crítica. 2013;26(2):241–50. doi:10.1590/S0102-79722013000200004.

Vessey J, Strout DT, DiFazio RL, Walker A. Measuring the youth bullying experience: A systematic review of the psychometric properties of available instruments. J Sch Health. 2014;84(12):819–43. doi:10.1111/josh.12210.

Volk AA, Camilleri JA, Dane AA, Marini ZA. Is adolescent bullying an evolutionary adaptation? Aggress Behav. 2012;38:223–38. doi:10.1002/ab.21418.

Webster-Stratton C, Reid MJ, Stoolmiller M. Preventing conduct problems and improving school readiness: evaluation of the incredible years teacher and child training programs in high-risk schools. J Child Psychol Psychiatry. 2008;49(5):471–88. doi:10.1111/j.1469-7610.2007.01861.x.