

2007

Item Wording and Internal Consistency of a Measure of Cohesion: The Group Environment Questionnaire

Mark A. Eys

Wilfrid Laurier University, meys@wlu.ca

Albert V. Carron

University of Western Ontario

Steven R. Bray

McMaster University

Lawrence R. Brawley

University of Saskatchewan

Follow this and additional works at: http://scholars.wlu.ca/kppe_faculty

Recommended Citation

Eys, Mark A.; Carron, Albert V.; Bray, Steven R.; and Brawley, Lawrence R., "Item Wording and Internal Consistency of a Measure of Cohesion: The Group Environment Questionnaire" (2007). *Kinesiology and Physical Education Faculty Publications*. 16.
http://scholars.wlu.ca/kppe_faculty/16

This Article is brought to you for free and open access by the Kinesiology and Physical Education at Scholars Commons @ Laurier. It has been accepted for inclusion in Kinesiology and Physical Education Faculty Publications by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact scholarscommons@wlu.ca.

Item Wording and Internal Consistency of a Measure of Cohesion: The Group Environment Questionnaire

**Mark A. Eys,¹ Albert V. Carron,² Steven R. Bray,³
and Lawrence R. Brawley⁴**

¹Laurentian University, ²University of Western Ontario,

³McMaster University, ⁴University of Saskatchewan

A common practice for counteracting response acquiescence in psychological measures has been to employ both negatively and positively worded items. However, previous research has highlighted that the reliability of measures can be affected by this practice (Spector, 1992). The purpose of the present study was to examine the effect that the presence of negatively worded items has on the internal reliability of the Group Environment Questionnaire (GEQ). Two samples ($N = 276$) were utilized, and participants were asked to complete the GEQ (original and revised) on separate occasions. Results demonstrated that the revised questionnaire (containing all positively worded items) had significantly higher Cronbach α values for three of the four dimensions of the GEQ. Implications, alternatives, and future directions are discussed.

Key Words: Group Environment Questionnaire, internal reliability, response acquiescence

In the field of sport psychology, one of the most widely used multidimensional measures of cohesion is the Group Environment Questionnaire (GEQ; Carron, Widmeyer, & Brawley, 1985). Indeed, in Dion's (2000) seminal review of cohesion concepts and measures, the GEQ features prominently as one of the more contemporary approaches to examining the many research questions associated with teams and groups in sport and physical activity. However, as noted by the developers of the GEQ (Brawley, Carron, & Widmeyer, 1987; Carron, Brawley, & Widmeyer, 2002), the refinement of a measure, including its reliability and validity, is an ongoing process. Accordingly, the present paper addresses one facet of this ongoing process—the internal consistency of the GEQ.

The GEQ evolved from the following definition of cohesion: “a dynamic process that is reflected in the tendency for a group to stick together and remain united in the pursuit of its instrumental objectives and/or for the satisfaction of member affective needs” (Carron et al., 1998, p. 213). More specifically, the GEQ

comprises subscales that assess the following constructs: (a) Individual Attractions to the Group–Task (ATG-T, the individual’s perceptions of his/her personal involvement in task aspects of the group), (b) Individual Attractions to the Group–Social (ATG-S, the individual’s perceptions of his/her involvement in social aspects of the group), (c) Group Integration–Task (GI-T, the individual’s perceptions of the degree of unity the group possesses surrounding task aspects), and (d) Group Integration–Social (GI-S, the individual’s perceptions of the degree of unity the group possesses regarding social aspects). Numerous empirical studies have supported the validity of the GEQ (for a review, see Carron et al., 1998).

More specific to the present study are issues related to the reliability of the GEQ. The most commonly used formulation to statistically assess scale reliability is the alpha coefficient, more generally referred to as *Cronbach’s alpha* (Cronbach, 1951). The issue of what is an acceptable value for the alpha coefficient was discussed by Pedhazur and Schmelkin (1991). They suggested that, generally, users should determine what amount of error they are willing to tolerate. However, they also recognized that researchers look to authoritative sources to justify standards of reliability. Currently, one authoritative source for the minimum standard of an alpha coefficient is Nunnally (1978), who suggested that an alpha value of .70 or higher is acceptable.

From the perspective of the GEQ, Carron and colleagues (1998) reported that the original Cronbach alpha values of the four scales were .75 (ATG-T), .64 (ATG-S), .70 (GI-T), and .76 (GI-S) and that “similar and larger values have been reported for the GEQ in other studies” (p. 222). This latter assertion notwithstanding, an examination of published and unpublished research carried out with the GEQ in a number of contexts yields two subsets of studies that reported internal consistency values. (A detailed table of references and previous internal consistency values of the GEQ examined for the present study can be obtained by contacting the lead author.) In the studies in one subset, the internal consistency values of the four scales have been shown to be acceptable according to the Nunnally recommendation (e.g., Li and Harmer, 1996; Carron et al., 1985). The second subset demonstrates that relatively variable internal consistencies have been reported on one or more of the scales within the GEQ. For example, Westre and Weiss (1991) computed internal consistencies that were below the .70 criteria (Nunnally, 1978) for each dimension (i.e., ATG-T = .68, GI-T = .66, ATG-S = .54, GI-S = .44). These results forced them to retain only the task scales in their study of coaching behaviors and group cohesion.

One aspect of original scale construction for the GEQ utilized by Carron and colleagues (1985) may be a contributor to variability in internal consistency, namely, the use of mixed stem (positive and negative) wordings for the various items. In general, the use of both negatively and positively worded items in a questionnaire was generated from research into a phenomenon termed *response acquiescence*. Also known as *agreement* (or *disagreement*) *tendency*, it is defined as “the tendency of an individual to agree or say ‘yes’ to . . . inventory statements, regardless of the content of the items” (Block, 1965, p. 1). Nunnally (1978) suggested that measures that employ a balance of negative and positive items could detect invalidating response sets. That is, response contradictions should be apparent if a respondent completes an inventory containing a balanced set of positive and negative items but agrees with all statements (i.e., engages in a process of response acquiescence).

A number of studies have investigated the effects of item wording on the validity and reliability of various inventories (e.g., Marsh, 1996). Researchers have commented on the effect of negative item wording on the internal consistency of a scale. Spector (1992), even though advocating the use of balanced scales, cautioned test developers to avoid the use of negation (i.e., adding *no* or *not*) in order to reverse the wording of an item. For example, the phrase “I am satisfied with my job” should not be transformed to “I am not satisfied with my job” but rather to “I hate my job.” Spector noted that there is the potential to misread the former statement and lead to a response on the wrong end of the scale. Although probably not having a major effect on the mean of a multiitem scale, this error has the potential to significantly reduce internal consistency. As more recently noted by Barnette (2000), items phrased using negation may not be “considered the exact opposite of directly worded items, and this is one of the major factors in the reduction of reliability and validity scores on surveys using mixed items” (p. 369). Barnette also concluded that research similar to his own should be undertaken with other measures.

It is clear that the use of the GEQ has, in some studies, been affected by lower scale reliability. It is also clear that the use of positively and negatively worded items can have an effect on the psychometric properties of an instrument. Thus, the purpose of our study was to examine whether positive wording of GEQ items would affect scale internal consistency. Related to this purpose, secondary issues were examined comparing the two measures as to the mean values and variances of their scale scores. However, due to space considerations these have not been presented. Results can be obtained by contacting the lead author.

Based on the response acquiescence and item valence research carried out with other instruments, it was hypothesized that Cronbach alpha values for GEQ scales using positively framed items would be higher than the alpha values for the original GEQ scales, which employ a mixed format (i.e., positive and negative items). Two samples of respondents were utilized. The first consisted of undergraduate students who were former team sport athletes. The second consisted of current members of intramural team sports.

Method

Participants

Undergraduate Sample. The participants were 195 undergraduate students from two Canadian universities (80 males, $M_{\text{age}} = 21.59 \pm 2.34$ years; 115 females, $M_{\text{age}} = 20.72 \pm 1.69$ years). The participants were current or former members of teams from both interactive (e.g., basketball; $n = 166$) and co-acting team sports (e.g., rowing; $n = 29$). The average tenure or association with teams was 3.41 years ($SD = 2.56$).

Intramural Sample. The participants were 81 intramural athletes from two Canadian universities (37 males, $M_{\text{age}} = 20.64 \pm 1.87$ years; 40 females, $M_{\text{age}} = 19.98 \pm 1.48$ years). Four participants did not divulge their gender. All participants were current members of interactive team sports at the intramural (i.e., recreational)

level participating in the sports of volleyball ($n = 7$), curling ($n = 4$), basketball ($n = 13$), and indoor soccer ($n = 57$). The average tenure with the team was 1.65 years ($SD = 1.66$).

Measures

Original Version of the GEQ. The 18-item GEQ (Carron et al., 2002) assesses four dimensions of cohesion: individual attractions to the group–social (ATG-S; five items); individual attractions to the group–task (ATG-T; four items); group integration–task (GI-T; five items); and group integration–social (GI-S; four items).

Participants respond to each of the 18 statements on a 9-point Likert scale anchored at 1 by *strongly disagree* and 9 by *strongly agree*. Thus, higher scores reflect stronger perceptions of cohesiveness. Twelve of the 18 items are negatively worded and are reverse scored (i.e., all four of the items in the ATG-T scale, three of the five items in the ATG-S scale, two of the five items in the GI-T scales, and three of the five items in the GI-S scale).

Modified Version of the GEQ. To assess the degree to which item wording affected the equivalency of the GEQ, it was necessary to modify the original GEQ so that all statements were phrased in a positive manner. Thus, for example, the original item “I do not enjoy being part of the social activities of this team” was modified to “I enjoy being part of the social activities of this team.” This process was carried out for each of the 12 items. Care was taken to avoid changing the meaning of any item in altering the language to a positive tone.

Procedure

Ethical approval was obtained through the lead author’s university ethics board. Participants completed *both* versions of the GEQ in the following manner. After consent was obtained, the participants first completed either the original GEQ (i.e., mixed format of positive and negative items) or the modified form (i.e., exclusively positively worded items). To counter potential order effects, half of the participants were randomized to receive the original version of the GEQ first whereas the remaining participants filled out the modified version first. To counter response set and similarity of question bias, all participants were given their second questionnaire to fill out 5 days after completing the first.

Results

Preliminary Analysis

To address the possible order effect, a within-subject 2 (order of presentation) \times 2 (GEQ version) repeated measures MANOVA was conducted for each of the undergraduate and intramural samples. For each sample, there was no multivariate within-subject interaction or main effect of order. Thus, the counterbalancing of GEQ versions to minimize order effects was successful.

Comparison of Equivalency Values (Internal Consistency)

As can be observed from Table 1, Cronbach α values for the positively worded questionnaire were higher across all GEQ scales. The statistical significance of these differences was evaluated using the following test statistic (see Feldt, Woodruff, & Salih, 1987; $df = N - 2$, where N = total number of participants):

$$t = (\alpha_1 - \alpha_2)(N - 2)^{1/2} \div [4(1 - \alpha_1)(1 - \alpha_2)(1 - \rho^2)]^{1/2}$$

where the α s are the internal consistency values from the two scales, and ρ represents the bivariate correlation between the two scales (Table 1).

The results for the *undergraduate sample* demonstrated that the alpha values for the positively worded questionnaire were significantly higher for the ATG-S, $t(193) = 6.99, p < .05$; GI-T, $t(193) = 4.19, p < .05$; and GI-S scales, $t(193) = 6.28, p < .05$. However, the difference between ATG-T scales was not significant, $t(193) = 1.06, p > .05$.

For the *intramural sample*, the results obtained were similar to the above in that the alpha values for the exclusively positively worded questionnaire were significantly higher for the ATG-S, $t(80) = 3.51, p < .05$; GI-T, $t(80) = 2.58, p < .05$; and GI-S scales, $t(80) = 3.94, p < .05$. Again, however, no significant difference was found in comparing the ATG-T scales for either GEQ version, $t(80) = 1.22, p > .05$.

Table 1 Alpha Values and Bivariate Correlations Between Cohesion Dimensions for the Undergraduate and Intramural Samples

Variable	α	1.	2.	3.	4.	5.	6.	7.	8.	α
1. Original ATG-S	.46	—	.48**	.45**	.65**	.75**	.34**	.29**	.60**	.73
2. Original ATG-T	.78	.52**	—	.55**	.41**	.25**	.57**	.38**	.20**	.66
3. Original GI-T	.73	.47**	.55**	—	.60**	.40**	.46**	.59**	.40**	.74
4. Original GI-S	.70	.57**	.31**	.42**	—	.59**	.33**	.36**	.70**	.83
5. Positive ATG-S	.74	.32**	.21	.47**	.50**	—	.45**	.39**	.74**	.86
6. Positive ATG-T	.83	.28*	.38**	.50**	.35**	.70**	—	.62**	.32**	.70
7. Positive GI-T	.84	.17	.20	.41**	.23*	.54**	.70**	—	.44**	.84
8. Positive GI-S	.86	.29**	.05	.31**	.49**	.69**	.57**	.70**	—	.91

Note. ATG-T = Individual Attractions to the Group-Task, ATG-S = Individual Attractions to the Group-Social, GI-T = Group Integration-Task, GI-S = Group Integration-Social. Original = original version of questionnaire, Positive = revised questionnaire. Correlations/alpha values for the undergraduate sample ($n = 195$) are presented in the upper-right portion of the table, whereas correlations/alpha values for intramural sample ($n = 81$) can be found in the lower-left portion.

* $p < .05$, ** $p < .01$.

Discussion

The purpose of our study was to examine whether positive wording of GEQ items would affect the internal consistency (equivalency) of the scales. There are a number of issues related to the purposes that warrant discussion. The first issue relates to the effect of item wording on the equivalency (internal consistency) of

the GEQ scales. In both of the samples examined, the positively worded version of the questionnaire had significantly higher Cronbach α values in three of the four dimensions. This finding appears to be in agreement with other research with other instruments; results showed that the combination of positively and negatively worded items significantly reduced a scale's internal reliability (Spector, 1992; Barnette, 2000).

The GEQ scale for which internal consistency values were similar (i.e., no significant difference in α values) for the two versions of the questionnaire was Individual Attractions to the Group–Task. Items making up the Attractions to the Group–Task subscale are worded exclusively negatively in the original GEQ; revision of those items resulted in exclusively positive wording. It may be that the uniformity of item valence on a scale (i.e., either all negative or all positive) would lead to similar internal consistency values when comparing versions. Our results are consistent with this interpretation; however, an examination of this issue would require future investigation.

Future investigation should also address an additional factor—the effect of individual differences. Specifically, there are individual differences among respondents with regard to the interpretation of *negative* items. For example, examining the cohesion of older adult physical activity groups, Estabrooks and Carron (2000) modified the sport-related GEQ by changing the wording to consider structured physical activity classes. Older individuals in these studies found negatively worded items difficult to interpret and expressed dissatisfaction at viewing their group in a negative fashion (i.e., a social reactivity bias). Thus, individual differences may also need to be explored in future investigation of positive vs. negative vs. mixed wording of items that represent perceptions of cohesion in sport.

Interestingly, our review of studies that have utilized the GEQ demonstrated that the dimension Attractions to the Group–Social has consistently exhibited less than acceptable alpha values. Although one possible explanation is the use of mixed format scales rather than positive format scales, Carron et al. (1998) highlighted another explanation. Specifically, they noted that the dimensions included in the GEQ are not necessarily present or salient at all points in time in a group's history. In addition, it is also possible that these dimensions will not be salient to the same degree *across* groups. For example, if one were to examine business or sport groups where the group's task orientation predominates in their interactions, social aspects of the group may not be a relevant perception of cohesion. It is possible that this was the case in the example studies that detected low internal consistency values for ATG-S.

Our exploratory study into the effect of wording on the equivalency of the GEQ has implications for both the strategies to combat response acquiescence as well the questionnaire itself. The current results support the contention that the practice of utilizing mixed items in the GEQ results in lower internal consistency scores than those obtained using a solely positive format even when acceptable alpha values are obtained through both formats.

For those concerned about response acquiescence, two strategies are offered based on suggestions by Barnette (2000). These strategies included balancing positively and negatively worded items but also switching the anchors for various questions. For example, on a 1-to-9 Likert item, this would mean labeling 1 as *Strongly Disagree* for some questions and *Strongly Agree* for others (and vice versa for 9).

However, this creates the possibility of confusing participants. Barnette warned his participants about this possibility before having them complete the questions. A simpler strategy may be to insert a general and spurious negatively worded item (e.g., for the GEQ, "My team is not cohesive") among relevant positively worded items to ensure subjects are not engaging in response acquiescence. If participants detect the change, then acquiescence would not be a consideration.

We would be remiss if we did not mention limitations of this initial study of the wording format of the GEQ. One is that only one variation of item wording was investigated. Future research could examine the viability of multiple strategies including mixed item wording as well as bidirectional response alternatives in relation to effects on the internal consistency of the GEQ scales. It should be noted that we investigated a strategy for improving the internal consistency of scales that for the most part had already acceptable values. The approach to improvement may benefit other measures with unacceptable values. Consequently, a scale that may have been discarded when utilizing the original version might now be included for additional analyses.

A second limitation is that we used two selective samples for this preliminary study. Even though a number of teams were represented in our undergraduate sample, we did not examine intact groups, a characteristic of numerous GEQ studies. For our results to be more generalizable, it would be useful to explore the effects of different wording formats on the internal consistency of randomly selected intact teams from different sports.

A third limitation pertains to the sole use of the alpha coefficient as the measure for scale reliability. There are problems related to alpha that are often overlooked in psychology and sport psychology literature, including the fact that this value is highly dependent on the number of items that make up the dimension of interest (Cortina, 1993). For example, Cortina suggested that examining the interitem correlation would provide additional information regarding the homogeneity of scale items. Clearly, however, the use of Cronbach alpha predominates in sport and exercise psychology literature and thus was the main interest in the present report.

It should be noted that attaining an appropriate or higher alpha level does not necessarily guarantee concomitant improvements in detecting effects in examining the relationship between cohesion and other variables. As noted by Carron, Brawley, and Widmeyer (2002) with regard to the predictive validity of the GEQ, it is not only the psychometrics of the GEQ that must be considered. It is also the reliability and validity of other measures that must be taken into consideration when investigating predictive relationships.

The GEQ continues to be the most widely used measure of cohesion to examine teams in the sport environment. The approach taken in the present study is consistent with the recommendations made by the developers of the GEQ in their guidelines for future use of the instrument. Specifically, they noted that the validation of an instrument is an ongoing process (Carron et al., 1998, 2002) and, because reliability places an upper limit on validity, empirical efforts to hone the measurement of cohesion have implications for the understanding of cohesion and its relationship with other variables.

Results of the present study should not be used to discount previous GEQ research. Instead, they highlight the possibility of continuing to refine the measure. As an example, consideration should be given to how individual differences (e.g.,

people of various ages and educational backgrounds) affect responses to different item wording formats for the GEQ items. Given that the questionnaire was derived from a strong theoretical framework and that “the validity of the GEQ has been repeatedly demonstrated in numerous studies” (Li & Harmer, 1996, p. 50), this measure and future improvements to it will continue to make it a valuable tool for the examination of group cohesion in sport.

References

- Barnette, J.J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement, 60*, 361-370.
- Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Meredith Publishing Company.
- Brawley, L.R., Carron, A.V., & Widmeyer, W.N. (1987). Assessing the cohesion of teams: Validity of the group environment questionnaire. *Journal of Sport Psychology, 9*, 275-294.
- Carron, A.V., Brawley, L.R., & Widmeyer, W.N. (1998). The measurement of cohesiveness in sport groups. In J.L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 213-226). Morgantown, WV: Fitness Information Technology.
- Carron, A.V., Brawley, L.R., & Widmeyer, W.N. (2002). *The Group Environment Questionnaire test manual*. Morgantown, WV: Fitness Information Technology.
- Carron, A.V., Widmeyer, W.N., & Brawley, L.R. (1985). The development of an instrument to assess cohesion in sport teams: The group environment questionnaire. *Journal of Sport Psychology, 7*, 244-266.
- Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Dion, K.R. (2000). Group cohesion: From “field of forces” to multidimensional construct. *Group Dynamics: Theory, Research and Practice, 4*, 7-26.
- Estabrooks, P.A., & Carron, A.V. (2000). The Physical Activity Group Environment Questionnaire: An instrument for the assessment of cohesion in exercise classes. *Group Dynamics: Theory, Research, and Practice, 4*, 230-243.
- Feldt, L.S., Woodruff, D.J., & Salih, F.A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93-103.
- Li, F., & Harmer, P. (1996). Confirmatory factor analysis of the group environment questionnaire with an intercollegiate sample. *Journal of Sport and Exercise Psychology, 18*, 49-63.
- Marsh, H.W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts. *Journal of Personality and Social Psychology, 70*, 810-819.
- Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Spector, P.E. (1992). *Summated rating scale construction: An introduction*. (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-082). Newbury Park, CA: Sage.
- Westre, K.R., & Weiss, M.R. (1991). The relationship between perceived coaching behaviors and group cohesion in high school football teams. *The Sport Psychologist, 5*, 41-54.

Manuscript submitted: June 27, 2006

Revision accepted: January 15, 2007