Journal of Applied Sport Psychology

ISSN: 1041-3200 (Print) 1533-1571 (Online) Journal homepage: https://www.tandfonline.com/loi/uasp20

How to Improve Athletes' Return of Investment: Shortening Questionnaires in the Applied Sport Psychology Setting

Stephan Horvath & Philipp Röthlin

To cite this article: Stephan Horvath & Philipp Röthlin (2018) How to Improve Athletes' Return of Investment: Shortening Questionnaires in the Applied Sport Psychology Setting, Journal of Applied Sport Psychology, 30:2, 241-248, DOI: 10.1080/10413200.2017.1382020

To link to this article: https://doi.org/10.1080/10413200.2017.1382020

Accepted author version posted online: 21 Sep 2017.
Published online: 18 Oct 2017.

JOURNAL OF APPLIED SPORT PSYCHOLOGY, 30: 241–248, 2018
ISSN: 1041-3200 print / 1533-1571 online
DOI: 10.1080/10413200.2017.1382020

Routledge
Taylor & Francis Group

RESEARCH NOTE

How to Improve Athletes' Return of Investment: Shortening Questionnaires in the Applied Sport Psychology Setting

Stephan Horvath and Philipp Röthlin

Swiss Federal Institute of Sport Magglingen

In counseling settings we need short and understandable, as well as valid and reliable, questionnaires. However, test construction focuses primarily on the latter, resulting in rather long tests and athletes' perception of low return on time investment. We address this problem by explaining how questionnaires can be shortened (i.e., using three selection criteria) and validated (e.g., by calculating interitem correlations). As an example we shortened the Thought Occurrence Questionnaire Sport from 17 to three items. We argue why this short version still satisfies quality criteria and why such short(ened) questionnaires should be more appreciated (e.g., by publishing and using them in counseling settings).

Having a background in basic psychology and working as applied researchers and sport psychology consultants, we sometimes feel that basic research and counseling work in sport psychology seem to take place in different worlds. This also holds for the use of questionnaires. Validated questionnaires often lack usability in a counseling setting because they are too long or unsuitable for the specific sport or athlete—for example, because they contain items that do not resonate with athletes. This is also the case when we collect research data in the (elite) sports context (Anshel & Brinthaupt, 2014). Therefore, we often must adapt validated questionnaires by shortening scales or reformulating items. The use of simplified instruments comes with some costs: First, it is hard to publish such data in research journals, because the instrument is no longer in its original version. Second, using small numbers of items to measure psychological concepts makes it difficult to satisfy quality criteria (e.g., Cronbach's alphas) used in basic research (Tavakol & Dennick, 2011). These findings are therefore rarely published and end up in researchers' drawers. We are convinced that these findings are worth being published and of interest to counseling psychologists or researchers. Thus, we—as applied researchers—should face these challenges of acceptance and take the necessary steps to offer refined psychometric procedures. As a result, our effort to improve sport psychological counseling may be better rewarded.

Received 14 June 2017; accepted 17 September 2017.

Address correspondence to Philipp Röthlin, Swiss Federal Institute of Sport, Alpenstrasse 16, CH-2532 Magglingen, Switzerland. E-mail: philipp.roethlin@baspo.admin.ch

In this research note, we address the problems and present possible solutions.[1] First, we outline that questionnaires in basic research and in the counseling setting have different functions. Next, we argue that adapted and/or shortened questionnaires should, of course, be validated to ensure validity and reliability. However, some of the criteria (i.e., major psychometric properties) should be weighted differently than in basic research (e.g., by focusing more on convergent, divergent, and predictive validity), and we should consider minor criteria (e.g., utility and fairness) as well. We also argue that we need short questionnaires in counseling settings to increase usability and demonstrate how such questionnaires can be developed and evaluated to ensure validity and reliability. In a final section, we describe how we adapted the Thought Occurrence Questionnaire Sport (TOQS; Hatzigeorgiadis & Biddle, 2000; German version, Röthlin, Horvath, Birrer, Güttinger, & grosse Holtforth, 2017) as a tool for our counseling work and applied research. We hope our research note brings applied research and short questionnaires more recognition and makes such research findings and instruments available to others. Counseling sport psychologists might benefit from applied research and their counseling-friendly, yet reliable and valid, instruments.

QUESTIONNAIRES IN SPORT PSYCHOLOGY: BASIC RESEARCH VERSUS COUNSELING SETTING

Questionnaires in basic research are important tools to measure psychological concepts and to test a researcher's hypotheses. Developing and translating questionnaires are important parts of basic research. This importance is highlighted by several journals focusing on measurement and assessment.

In the counseling setting observations, conversations with the athlete and related coaches and, potentially, the use of questionnaires are methods that provide information to get a complete picture about an athlete's situation. We see two main reasons why questionnaires are of aid. First, they can help sport psychologists screen for topics that should be addressed and might be worth working on, such as when they want to know whether specific factors negatively affect an athlete's performance or development (e.g., performance anxiety). Second, questionnaires can help monitor important changes (e.g., current stress and recovery). This shows that there is a need for questionnaires in counseling psychology. It may, therefore, come as a surprise that questionnaires developed in the basic research field are not regularly used in the counseling setting. Sometimes, they are replaced by self-generated, nonvalidated surveys (O'Connor, 2004). This indicates that basic research does not develop questionnaires that are suitable for counseling settings.

There are a number of reasons why "original" questionnaires developed in basic research are not always used for counseling purposes. One important factor is participants. Counseling sport psychologists work mainly with elite athletes or with young athletes. The former have full schedules and prefer spending their time exercising than completing questionnaires. For the latter, it is important that they understand the questions; thus, they must be formulated accordingly (i.e., simple, and they must make sense in the specific sport context). Athletes should generally see some benefit in completing a questionnaire, and as a prerequisite, questionnaires should be short and clear. In sum, counseling sport psychologists need questionnaires that are short (i.e., one to four items to measure a concept) and easy to understand. In contrast, basic research produces questionnaires that are often long and include similar items, some of which are hard to understand and not suitable for all kind of sports. The rather high number of items may be directly related to the goals that questionnaires in basic research must

fulfill, for example, capturing a concept as exactly as possible. In sum, basic research produces questionnaires for basic research and journals but not for everyday use in sport psychological counseling.

We conclude that counseling sport psychologists need short questionnaires that can be used for screening and as monitoring tools. Basic research, on the other hand, develops long scales to satisfy scientific credibility. This comes at the expense of usability in everyday counseling. If adapted questionnaires are used, the data collected hardly ever get published and made available to other researchers. We think that this gap between the two fields should be filled by a third field, namely, applied research. We need short questionnaires with high usability that remain valid and reliable (for problems of short forms, see Smith, McCarthy, & Anderson, 2000). In the following section, we explain how this could be done.

FILLING THE GAP BETWEEN SPORT PSYCHOLOGICAL RESEARCH AND SPORT PSYCHOLOGICAL COUNSELING: APPLIED RESEARCH

Basic research pursues a contribution to our theoretical knowledge in a field. Therefore, researchers must fulfill high standards. When developing questionnaires, these standards include major psychometric properties, that is, reliability (e.g., retest reliability, internal stability, or interrater stability) and validity (e.g., content validity, construct validity, convergent/divergent validity, or predictive/criterion validity), and objectivity (e.g., when collecting, analyzing, and interpreting individual data; e.g., Furr & Bacharach, 2013). The more items used to measure a construct, the easier it is to satisfy reliability (Tavakol & Dennick, 2011). Although reliability and validity are important to estimate the quality of an instrument, we think that in applied research, not all traditional indicators of reliability and validity are as important as in basic research. On the other hand, less prominent psychometric properties (e.g., fairness) become more relevant, although they often cannot be quantified. Therefore, we want to discuss appropriate and inappropriate indicators of psychometric properties when evaluating short questionnaires used for counseling or in applied research in this section. Besides reliability, validity, and objectivity, we want to call attention to some minor psychometric properties that should be weighted more when evaluating the quality of short questionnaires in the counseling setting.

Major Psychometric Properties

Reliability

High reliability indicates that a questionnaire measures the construct of interest in a stable manner. The most common indicator for gauging reliability in a questionnaire is internal consistency (Cronbach's alpha). Alpha is calculated based on correlations between all items of a questionnaire, as well as the number of items. Alphas range between zero and one. Tovakol and Dennick (2011) mentioned that there is no consensus about the threshold for acceptable alpha scores; they range from 0.70 to 0.95. Higher reliabilities are not desirable, as this indicates that some items may be entirely redundant and that the test length should be shortened (Tavakol & Dennick, 2011). Unfortunately, number of items affects Cronbach's alpha. Actually, short questionnaires produce lower alphas, even if they have the same average interitem correlations. Therefore, we recommend reporting interitem correlations when using or developing (short) questionnaires. According to Clark and Watson (1995), the average interitem correlation for a narrow construct should be between .40 and .50. Stable concepts can also be validated using retest reliabilities.

Validity

Every questionnaire, no matter its number of items, must be valid, that is, it has to measure the concept of interest. Most indicators of validity are applicable to short questionnaires. Content validity asks whether the items represent the definition of a concept. That is a challenge when dropping items to abbreviate a questionnaire. Statistical analysis (viz., correlations) are often used to examine content validity. Shortened questionnaire should show the same convergent and divergent validity as the original version. Expected high correlations with other constructs indicate convergent validity, whereas expected low or zero correlations indicate divergent validity. Finally, in the counseling setting, predictive/criterion validity is an important indicator (Tenenbaum, Eklund, & Kamata, 2012). We are looking for a (shortened) questionnaire that can predict connected outcomes. For example, if we want to know whether an athlete is choking under pressure, a good choking questionnaire is one that can predict choking in the near future.

Objectivity

The last major psychometric property is objectivity. A questionnaire should be used, analyzed, and interpreted the same way, no matter who uses the test. If a good test manual is available, the length of the questionnaire should not affect objectivity. However, it might be easier to follow the manual if the questionnaire is short.

Aside from these major psychometric properties, there are some minor psychometric properties that should be considered, especially when evaluating questionnaires in a counseling setting (Moosbrugger & Kelava, 2007). These minor criteria might have been neglected because, in contrast to major psychometric properties, they are harder to quantify.

Some Minor Psychometric Properties

Whether the questionnaire is applicable for a wider population in real-world situations (especially counseling or sport settings) is a question of *utility*. As we showed before, questionnaires developed in the basic research field are often of little value for counseling sport psychologists' purposes. *Test economy* asks for the effort–benefit ratio. Counseling sport psychologists are looking for questionnaires with a good return on investment. A higher number of items can increase reliability and validity but worsens the effort–benefit ratio. The questionnaire must be *fair*, especially if we want to compare athletes. Fairness in this context includes the idea that everyone understands the questions the same way and that individual differences in cognitive abilities should not affect the answers. Thus, fairness asks for clear and short questionnaires. Whereas test economy focuses on the investment of the researcher and the value of the outcome, *reasonableness* focuses on the investment of the athlete. As sport psychological consultants, we are service providers, and the athletes are our clients. Their time is limited, and the results of questionnaires are only a first step in consultations. Afterward, the results are discussed and verified, and if necessary an intervention is initiated. In such a setting, completing questionnaires should take as little time as possible.

In sum, we suggest that applied research should weight psychometric properties differently than basic research. In applied research, a questionnaire should primarily satisfy reliability (i.e., high interitem correlation), predictive/criterion validity, but weigh utility and reasonableness as equally important. This comes with some shortcomings but increases the questionnaire's applicability in consultations.

HOW TO ADAPT LONG QUESTIONNAIRES FOR APPLIED RESEARCH AND COUNSELING?

The previous sections showed that questionnaires developed in basic research satisfy major psychometric properties but that they are often too time consuming for the athletes. This may produce a higher percentage of missing values (Alcaraz, Viladrich, & Torregrosa, 2013). In addition some items may not make sense in a specific sport context or for some athletes. Thus, these questionnaires should be adapted before they are used in applied research and counseling. As we mentioned earlier, some sport psychology consultants use self-created questionnaires (O'Connor, 2004) but unfortunately do not validate them (Smith et al., 2000). We strongly advise against using such questionnaires. Convergent, divergent, and predictive/criterion validity can and should always be examined before a (short) questionnaire is used, independent of the number of items used to measure a concept. When shortening a questionnaire, we must identify the items that best represent the concept and decide how many items are sufficient to measure it.

The number of items needed depends on the construct. Well-known single-facet concepts such as self-esteem can even be measured with a single item (Robins, Hendin, & Trzesniewski, 2001). Because single-item questionnaires may lead to a small variance among participants, we recommend using three items and wide scales (i.e., 7-point Likert scales or 100-point visual analogue scales, depending on the measured concept).

There are three important criteria when choosing items. First, every item should represent the concept of interest and differentiate between high and low values on that particular dimension, that is, each item should have a high discrimination coefficient. Second, the chosen items should show different difficulties, that is, the percentage of people who agree with an item should be different for the selected items to increase variance among participants. Finally, and equally important, the selected items should be easy to understand for all participants. Comprehensibility can be evaluated by applied researchers or, even better, athletes—especially younger ones, as they might be prone to have the most problems in understanding an item. Next we describe how we shortened a 17-item questionnaire.

Other than choosing a subset of items and dropping the rest, chunking items into a single question is an alternative. For example, Hitzschke and colleagues (2016) shortened their Acute Recovery-Stress Questionnaire with eight subscales from 32 items to eight items. In the short form, participants are asked whether they feel emotionally balanced (e.g., satisfied, in balance, in a good mood, under control), whereas in the long version, each of these four items had to be rated separately. A similar strategy is to describe a concept in more detail, based on the different items used in the original version, before asking participants to rate one single item (e.g., Konrath, Meier, & Bushman, 2014).

EXAMPLE: MAKING A SHORT VERSION OF THE THOUGHT OCCURRENCE QUESTIONNAIRE SPORT

Cognitive interference means a disruption in concentration by thoughts in a performance situation. The TOQS (Hatzigeorgiadis & Biddle, 2000) measures cognitive interference with 17 items. We translated the TOQS into German and validated the instrument (Röthlin et al., 2017). In this last part, we describe how we shortened the TOQS to a three-item instrument and present some indicators that underline its reliability and validity.

Choosing the Items

Our goal was to construct a short instrument that covers the main concept of the TOQS, namely, cognitive interference. We selected three items based on three criteria: discrimina-

Table 1
Items of the TOQS, Selection Criteria and the Three Items Selected for the Short Version

Item no.	"During the competition/game I had thoughts …"	Corrected item-scale correlation	Item *M*
1	… that I want to quit.	0.49	1.73
2	… about other activities (e.g., shopping, having tea, TV).	0.44	1.82
3	… about previous mistakes I have made.	0.44	3.24
4	… *that I do not want to take part in this competition anymore.*	**0.66**	1.78
5	… about what I'm going to do later in the day.	0.47	2.16
6	… that I'm having a bad day.	**0.59**	2.62
7	… that I want to get out of here.	**0.64**	1.74
8	… about personal worries (e.g., school, work, relations).	0.51	2.02
9	… that the conditions (weather, temperature, pitch, atmosphere) are no good.	0.40	2.70
10	… about stopping.	0.53	1.63
11	… about friends.	0.33	2.42
12	… *that I am not going to achieve my goals today.*	**0.64**	2.80
13	… that I am fed-up with it.	**0.66**	1.86
14	… about what I'm going to do when I'll go home.	0.53	2.02
15	… that I am not going to win this competition.	**0.61**	3.05
16	… that I cannot stand it anymore.	0.52	1.42
17	… *that other competitors are better than me.*	**0.67**	3.17

Note. Bold items represent the seven items that show the highest item-scale correlation (first selection criterion); the three selected items for the short version are in italics. Items were rated on a 7-point Likert scale.

tion coefficient, item difficulty, and comprehensibility (Tenenbaum et al., 2012). The selection and validation are based on the data of the same 132 participants described by Röthlin and colleagues (2017). To identify how well a single item represents cognitive interference (TOQS total score), we calculated the corrected item-scale correlations (correlation between the single item and the mean of the remaining 16 items, i.e., the discrimination coefficient). The higher the correlation, the better an item represents the general concept. We preselected seven items based on their discrimination coefficient. In Table 1, the highest correlations are in bold print. To get a wide spectrum of scores—and differentiation between people—the chosen items should have different means (i.e., item difficulty). The preselected items can be grouped into low (1.74–1.86), middle (2.62–2.80), and high (3.05–3.17) mean scores (i.e., difficulty). We chose one item of each level applying the criterion of comprehensibility. It is important to choose items that are understood by all participants and in all sport contexts. For example, we selected Item 4 instead of Items 7 or 13 from the low-difficulty level (see Table 1).

VALIDATING THE SHORT VERSION

Statistical indicators for reliability and validity are presented in Table 2. Due to the smaller number of items, Cronbach's alpha of the three-item short version of the TOQS was lower than for the 17-item scale but still good. On the other hand, the mean interitem correlation of the short TOQS was good and due to the selection process higher than for the original version of the TOQS. Finally, retest reliability was similar for both versions of the TOQS. A first indicator for good validity is the high corrected correlations of $r = .78$ between the original TOQS (without the three items selected for the short version) and the short version. Furthermore, the short version shows similar convergent and divergent validities as the original TOQS (e.g.,

Table 2
Descriptives, Reliabilities, Convergent, Divergent, and Criterion Validity of the Original TOQS and the Short Version

	TOQS Original	TOQS Short
M	2.24	2.59
SD	0.74	1.08
Cronbach's α	.89	.76
M interitem correlation	.32	.51
r_{tt}[a]	.82**	.71**
Concentration disruptions	.49**	.46**
Cognitive anxiety	.53**	.63**
Somatic anxiety	.29**	.34**
Rumination	.26**	.32**
Mindfulness (total)	−.46**	−.50**
Acting with awareness	−.46**	−.47**
Decentering and nonreactivity	−.34**	−.38**
Accepting and nonjudgmental orientation	−.25**	−.34**
Openness to experience	−.06	−.09
Performance under pressure	−.40**	−.45**

Note. r_{tt} = Retest-Reliability (4 months). TOQS = Thought Occurrence Questionnaire Sport.
[a]Subsample ($n = 67$).
**$p < .01$.

cognitive anxiety, rumination, or mindfulness; Röthlin et al., 2017). In sum, our three-item TOQS seems to be a short, but still reliable and valid, tool to measure cognitive interference in the sport context. Furthermore, the short version might have better minor psychometric properties (viz., higher utility, test economy, fairness, and reasonableness) than the "original" version.

FINAL THOUGHTS

We argued why in the counseling setting we need other questionnaires than the ones produced in basic research. This gap should be filled by applied research. The validation of short questionnaires requires a reweighting of psychometric properties. We demonstrated exemplarily the different steps of the selection process and how the shortened version can be compared with the original version. The main goal of our example was to illustrate a simple way of shortening a questionnaire. However, the validation part has one important shortcoming. We have not examined the short version as a stand-alone instrument. Instead, participants completed the original long version of the questionnaire, and the short version's score was calculated afterward. Thus, responses to the three selected items might have been influenced by the other not-selected items. Accordingly, to confirm and strengthen the quality of shortened instruments, their validity and reliability should be examined with a new sample.

Short questionnaires are needed in the counseling setting, and we hope that their development is more appreciated and published in the future. This also should help increase the quality and economy of sport psychological consultations by adding an important piece of information about the athletes. Finally, we hope that these short forms become an inherent part of everyday sport-psychological counseling.

FOOTNOTE

[1]We focus on how to shorten one-dimensional concepts including total scores. Multidimensional concepts bare specific challenges that we do not address in this research note.

ORCID

Philipp Röthlin http://orcid.org/0000-0003-2268-571X

REFERENCES

Alcaraz, S., Viladrich, C., & Torregrosa, M. (2013). Less time, better quality. Shortening questionnaires to assess team environment and goal orientation. *The Spanish Journal of Psychology*, *16*, E77. doi:10.1017/sjp.2013.74

Anshel, M. H., & Brinthaupt, T. M. (2014). Best practices for the use of inventories in sport psychology consulting. *Journal of Clinical Sport Psychology*, *8*, 400–420. doi:10.1123/jcsp.2014-0045

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319. doi:10.1037/1040-3590.7.3.309

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.

Hatzigeorgiadis, A., & Biddle, S. J. H. (2000). Assessing cognitive interference in sport: Development of the thought occurrence questionnaire for sport. *Anxiety, Stress, and Coping*, *13*, 65–86. doi:10.1080/10615800008248334

Hitzschke, B., Holst, T., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2016). Entwicklung des Akutmaßes zur Erfassung von Erholung und Beanspruchung im Sport [Development of the Acute Recovery and Stress Scale]. *Diagnostica*, *62*, 212–226. doi:10.1026/0012-1924/a000155

Konrath, S., Meier, B. P., & Bushman, B. J. (2014). Development and validation of the single item narcissism scale (SINS). *PLoS One*, *9*(8), e103469. doi:10.1371/journal.pone.0103469

Moosbrugger, H., & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion [Test theory and Scale construction]*. New York, NY: Springer.

O'Connor, E. A. (2004). Which questionnaire? Assessment practices of sport psychology consultants. *The Sport Psychologist*, *18*, 464–468. doi:10.1123/tsp.18.4.464

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*, 151–161. doi:10.1177/0146167201272002

Röthlin, P., Horvath, S., Birrer, D., Güttinger, L., & grosse Holtforth, M. (2017). Kognitive interferenz im sport—validierung einer deutschsprachigen version des "Thoughts Occurence Questionnaires Sport" (TOQS) [Cognitive interference in sports–validation of a German version of the "Thoughts Occurence Questionnaire Sport" (TOQS)]. *Diagnostica*, *63*, 112–121. doi:10.1026/0012-1924/a000167

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

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. doi:10.5116/ijme.4dfb.8dfd

Tenenbaum, G., Eklund, C. R., & Kamata, A. E. (2012). *Measurement in sport and exercise psychology*. Leeds, UK: Human Kinetics.