

Adaptation and validation of the Polish version of the team boosting behaviors scale

Polish
adaptation of
the TBB scale

Joanna Haffer

WSB Merito University in Torun, Torun, Poland

Received 9 November 2022
Revised 18 April 2023
4 October 2023
19 January 2024
Accepted 23 February 2024

Abstract

Purpose – The article aims to present the results of adapting the team boosting behaviors (TBB) scale to Polish cultural conditions and validating it.

Design/methodology/approach – The research methodology consisted of three steps. In the first step, I translated the TBB scale into Polish using a rigorous back-translation method. Next, to assess content validity, nine domain experts reviewed the initial version of the instrument for clarity and relevance. Finally, I applied the scale to a sample of 532 team members and underwent thorough psychometric testing to assess construct validity. I employed structural equation modeling (SEM) with the partial least squares (PLS) factor-based algorithm technique for confirmatory factor analysis to assess the scale's reliability and validity.

Findings – After development, the Polish version of the TBB scale kept its three sub-scale structures. However, the validation process led to a slight reduction in the number of test items compared to the original scale.

Research limitations/implications – The findings imply that the Polish version of the scale is a valid and reliable tool for assessing TBB. However, I recommend additional studies to confirm this instrument's structure.

Originality/value – The results confirmed the reliability and relevance of the tool for measuring TBBs in Polish cultural conditions. The tool provides the basis for implementing further research with the TBB construct in Poland and internationally.

Keywords Content validity, Construct validity, Scale adaptation, Team boosting behaviors, Psychometric properties

Paper type Research paper

Introduction

Teamwork, which is required for effective team performance (Salas, Shuffler, Thayer, Bedwell, & Lazzara, 2015), is a dynamic, adaptive and episodic process that considers team members' feelings, thoughts and behaviors as they work together to achieve a common objective (Salas *et al.*, 2015). Effective teamwork depends heavily on the individual contributions of team members. However, the contributions distribution is not always equal

JEL Classification — M12

© Joanna Haffer. Published in *Central European Management Journal*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

This research was funded in whole by National Science Centre, Poland (No: 2021/05/X/HS4/00130).

I am grateful to Dr Heleen van Mierlo of Erasmus University Rotterdam for her kind consultation and valuable substantive comments received during the research project. I would also like to thank the participants of the expert panel for their commitment and time and the reviewers of my article for their valuable input.

Declaration of conflicting interests: The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.



Central European Management
Journal
Emerald Publishing Limited
e-ISSN: 2658-2430
p-ISSN: 2658-0845
DOI 10.1108/CEMJ-11-2022-0194

among all team members, as each person brings their own unique personality and position to the team, which affects its functioning (Mathieu, Maynard, Rapp, & Gilson, 2008; Mathieu, Tannenbaum, Donsbach, & Alliger, 2014). While some individuals can be extremely influential, one toxic team member can cause group dysfunction (Felps, Mitchell, & Byington, 2006). Nevertheless, a single team member can also strengthen the team or even make it successful (Fortuin, van Mierlo, Bakker, Petrou, & Demerouti, 2021). The literature uses the term “bad apples” for individuals whose dysfunctional behaviors negatively impact a team (Felps *et al.*, 2006). They may disrupt the team through, e.g. their negative attitudes or behaviors that negatively affect others and even block key group processes. Felps *et al.* (2006) identified the following three categories of challenging team member behaviors: effort withholding, negative emotional attitudes and violation of important interpersonal norms. When unchecked, these can carry negative consequences for the team. In contrast, the positive equivalent of the “bad apple,” and therefore its opposite, is a person who strengthens the team and its members’ spirit (Fortuin *et al.*, 2021). The informal term “the life of the party” has sparked research suggesting a new concept to understand the behaviors of such individuals. In general, the phrase describes individuals who feel comfortable among people and attract others. Such people are able to “light up a room,” enliven the atmosphere and foster an inspiring and positive environment. In a set of three studies that combined qualitative and quantitative techniques, D.J. Fortuin, H. van Mierlo, A.B. Bakker, P. Petrou and E. Demerouti developed the new concept of team boosting behaviors (TBBs) and created and validated a scale to measure them (Fortuin *et al.*, 2021).

Preliminary research has shown that the TBB scale holds promise for research, but further empirical studies are necessary to verify it (Fortuin *et al.*, 2021). Moreover, the literature does not yet include studies on TBB that take into account Polish cultural conditions. High-quality research on organizational management science that enables evidence-based theory development demands the use of carefully constructed and validated research tools (Heggestad, Scheaf, Banks, Hausfeld, Tonidandel, & Williams, 2019), especially in terms of measurement constructs such as TBBs that are expressed at a high level of abstraction. Furthermore, there is a strong need for research tools that are cross-culturally validated due to global population diversity (Sousa & Rojjanasrirat, 2011). Among the important benefits of cross-cultural research in the field of organizational behavior is that it can help reveal the universal and specific organizational phenomena. Moreover, it can expand the variability range in the phenomena under study (Aycan & Gelfand, 2012). By expanding the scope of variation, cross-cultural research can expand (and improve) theories and identify neglected dimensions that are critical in a given cultural context (Aycan & Gelfand, 2012). The role of cross-cultural research is crucial, especially when behaviors based on perceptual, cognitive, or personality bases are the subject of study and should increase even more when behaviors are studied in social and organizational contexts (Lonner & Malpass, 1994). Thus, above all, I aimed to adapt and validate the Polish language version of the TBB scale. Translation, adaptation and validation of the scale for cross-cultural research require careful planning and the use of rigorous methodological approaches to generate a valid and reliable tool for assessing the phenomenon of interest among the intended audience (Gudmundsson, 2009). In the following sections of this article, I will present the theoretical background, including the concept of TBB and the expected results of cross-cultural validation of the scale. Then, I will describe in detail each of the three research stages. In the first stage, I translated the TBB scale from English into Polish. In the second stage, a panel of experts assessed its content relevance. Finally, in the third stage, I presented the full psychometric features of the Polish version of the scale for consideration of its use in an applied context.

Theoretical background

The TBBs concept is consistent with the developing fields of Positive Organizational Behavior (e.g. Luthans, 2002; Wright, 2003) and Positive Organizational Scholarship (POS) (e.g. Cameron, Dutton, & Quinn, 2003; Cameron & Caza, 2004). In particular, in the field of organizational behavior, scholars emphasize the need and importance of a proactive, positive approach that focuses on strengths rather than continuing the spiral of negativity (Luthans, 2002). Similarly, POS aims to direct the attention of researchers and management practitioners towards the positive aspects in organizations that are associated with employees' positive emotions, positive group dynamics like forming good relationships and trust (Glińska-Noweś, 2017) and fostering creative processes (Haffer & Glińska-Noweś, 2013). Moreover, TBBs are related to team functioning and performance in an overly positive way through the interaction of various mechanisms (Fortuin *et al.*, 2021). They can start a positive affect spiral that results in a number of advantageous things, such as greater coordination and cooperation, diminished interpersonal conflict, stronger team performance, enhanced personal well-being and decreased absenteeism and turnover (Walter & Bruch, 2008; Collins, Lawrence, Troth, & Jordan, 2013). In Fortuin's study, TBBs were linked to positive individual and team phenomena, giving a hint of the potential positive implication of these behaviors for team performance (Fortuin *et al.*, 2021).

TBBs exemplify individual interpersonal behaviors in teams characterized by dominance and energy, positive expression and social attitudes (Fortuin *et al.*, 2021). Dominant behaviors exude assertiveness and energy. Positive expression refers to behaviors that are impulsive, playful and focused more on the team than on effective task performance (Fortuin *et al.*, 2021). On the other hand, social focus is expressed in cordial, sociable interpersonal behaviors designed to connect team members. All team members may exhibit these behaviors to varying degrees (Fortuin *et al.*, 2021). Looking at it within the organization, TBBs can be nurtured and coached. With support from managers or organizational practices, they can positively influence the entire team, starting from the lower levels. (Fortuin *et al.*, 2021).

The TBB scale is a self-reporting instrument that consists of 18 items to measure TBBs on the following three sub-scales: energizing behaviors, mood-enhancing behaviors and uniting behaviors (Fortuin *et al.*, 2021). Energizing behaviors involve, for example, coming up with new ideas and initiatives for organizing and participating in team activities, inventing games or starting friendly rivalries during difficult moments and being agents of change or innovation and thus taking a dominant/assertive position. Mood-enhancing behaviors include, e.g. telling funny stories and presenting negative team events in a positive light or even turning them into positive events. Unifying behaviors aim at other team members and involve making connections and building relationships between team members, which may occur, e.g. through involving the entire team in joint activities or engaging each team member in casual conversations and asking them about their jobs, interests and personal lives. Fortuin *et al.*'s (2021) results confirmed the factorial, convergent and criterion validity and reliability of the three-dimensional construct.

Originally, scholars validated the TBB scale in Dutch cultural conditions. However, it may require adaptation to other cultural conditions. Hofstede was the pioneer of cultural values research. He developed a model consisting of five dimensions, which serve to determine the impact of culture on companies. The investigated cultural dimensions are power distance, individualism, masculinity, uncertainty avoidance and long-term orientation (Hofstede, 1980; Hofstede, Hofstede, & Minkov, 1991). Hofstede's cultural dimensions offer a chance to highlight the similarities and variations between countries. Such a comparison (Country Comparison Tool, 2023) shows large differences in individual cultural dimensions for the Netherlands (NL) and for Poland (PL) (power distance: NL = 38, PL = 68; individualism: NL = 80, PL = 60; masculinity: NL = 14, PL = 64, uncertainty avoidance: NL = 53, PL = 93, long-term orientation: NL = 67, PL = 38). These data indicate the legitimacy of validating the

original Dutch scale on the ground of Polish culture, as the two countries differ culturally in many aspects. Moreover, these differences may have clear implications for teamwork and affect team processes and team dynamics, because many team phenomena are culture-specific (Gibson, 1999). For example, a low level of power distance promotes direct and honest communication and facilitates teamwork (Schneider & Barsoux, 2003), collectivists are more likely than individualists to value harmony in groups and find teamwork to be more fulfilling than working alone (Kirkman, Gibson, & Shapiro, 2001). Hence, I expect that, e.g. a significantly higher value of the power distance dimension for Polish culture may lead to differences in the area of mood-enhancing behaviors (the second dimension of the TBB construct), while a higher level of the masculinity dimension may lead to differences in the area of uniting behaviors (the third dimension of the TBB construct). The following detailed description of the three research stages will help verify this expectation.

Stage 1: translation

In the translation process, I followed the procedure recommended by Hambleton (2005) and I used the back-translation method along several steps and involved professional translators. Two independent professional translators (translator 1 and translator 2), both native Polish speakers, translated the research tool's instructions, items and response format in its original language into Polish. This resulted in two forward-translated versions of the research tool (PL1 and PL2). A third independent translator (translator 3) compared the two forward-translated versions of the research tool (PL1 and PL2) and then compared both of them with the original research tool. This process resulted in the initial, first translation of the instrument into Polish. Next, bilingual translators 4 and 5 back-translated the first translation of the instrument in Polish into the source language. Both translators 4 and 5 were fluent in English and Polish and they did not see the original scale. Translators developed two back-translated editions of the research tool (B-PL1 and B-PL2). A committee of three translators (translator 4, translator 5 and translator 3) discussed and settled any inconsistencies or ambiguities regarding the semantic, idiomatic, experiential and conceptual equivalence of the instructions, items and response format between the two back-translations (B-PL1 and B-PL2) and among each of the two back translations (B-PL1 and B-PL2) and the original research tool to deliver the final instrument in Polish (Beaton, Bombardier, Guillemin, & Ferraz, 2000). All translators involved had many years of experience translating texts in the field of psychology and management.

Stage 2: expert panel: content validity assessment

For further instrument examination and to establish content validity (Lynn, 1986), I adapted and applied the exclusive step proposed by DeVellis (2016) for scale development in which items are reviewed by experts. In line with Rubio, Berg-Weger, Tebb, Lee and Rauch (2003), "content validity refers to the extent to which the items on a measure assess the same content or how well the content material was sampled in the measure" (p. 94). There is widespread agreement among researchers about defining content validity and the methodological approach that should be taken to do so (Polit & Beck, 2006). Assessing the scale's content validity is a crucial first step in improving an instrument's construct validity. Hence, it constitutes an important topic for researchers who need high-quality measurements (Haynes, Richard, & Kubany, 1995). In this study, I examined content validity based on the responses of experts classified as people who worked in teams, either as researchers or in practice as practitioners. It is recommended that a minimum of three experts be consulted when assessing content validity (Lynn, 1986). Nevertheless, the maximum number of experts involved in the process has not been defined in the literature and employing more than 10

experts in the process is of doubtful use because as the number of experts increases, the probability of reaching a consensus decreases (Lynn, 1986; Polit & Beck, 2006). Therefore, I selected nine domain experts, namely three team leaders, three team members and three experts who had a reputation for being team boosters. I conducted all interviews in person using a questionnaire including a guide. At the outset of each interview, I stated its purpose, defined the TBBs and their individual dimensions and provided examples of these behaviors. I asked the expert panel to make a professional subjective judgment on the instructions, each item and the response format of the instrument according to their relevance to the construct and their clarity. The analysis of the responses to these questions took a quantitative form due to the questions scaling. I used the content validity index (CVI) and the modified Kappa statistic, an index that considers chance agreement, to gauge the viewpoints of the domain experts. Moreover, I tasked the experts with determining whether the items contained and covered all relevant details or if any were absent. Experts could also provide feedback and comment on each item. The questions in this regard were open-ended. I subjected the obtained responses to qualitative analysis through manual coding.

Content validity index

For a new or revised scale, the CVI is a reliable and extensively used method for assessing content validity (Polit, Beck, & Owen, 2007; Shrotryia & Dhanda, 2019), and it provides information about “the degree to which an instrument has an appropriate sample of items for construct being measured” (Polit & Beck, 2006, p. 493). The experts evaluated the CVI on two four-point scales (regarding clarity: “1 – unclear, 2 – somewhat unclear, 3 – quite clear, 4 – clear; regarding relevancy: 1 – not relevant, 2 – somewhat relevant, 3 – quite relevant, 4 – highly relevant”) (e.g. Davis, 1992; Zamanzadeh, Rassouli, Abbaszadeh, Majd, Nikanfar, & Ghahramanian, 2014). I employed the four-point scale so as not to use a middle score that could be both neutral and ambiguous (Lynn, 1986). According to Lynn (1986), CVI values can be calculated for each item on a scale (item-level content validity index, I-CVI) and the overall scale (scale-level content validity index, S-CVI). I calculated I-CVI as the number of experts providing a score of 3 or 4 divided by the total number of experts (the proportion of agreement regarding clarity and relevancy) (Polit & Beck, 2006). When there are more than five experts, the I-CVI should not be lower than 0.78 (Lynn, 1986; Polit & Beck, 2006). We may calculate S-CVI using two alternative methods. The first, i.e. the universal agreement calculation method (S-CVI/UA), requires consensus among all experts. It is defined as the proportion of items on an instrument that received ratings of 3 or 4 by all experts. The second, the averaging calculation method (S-CVI/Ave), is the average proportion of items granted 3 or 4 by all of the judges (Lynn, 1986; Polit & Beck, 2006; Waltz, Strickland, & Lenz, 2005). Although the averaging approach is less conservative and liberal in interpretation, it is still preferred, especially when the validation panel includes many experts (Polit & Beck, 2006). There are three methods for calculating the S-CVI/Ave. However, it is most advisable to calculate it using the average I-CVI (Polit & Beck, 2006). To indicate content validity, scholars recommend a minimum S-CVI/Ave of 0.9 (Waltz *et al.*, 2005; Polit & Beck, 2006). Items that do not meet the minimum allowable indices are re-evaluated. Concluding, to provide evidence for excellent content validity, the scale should include items with I-CVIs that fulfill Lynn’s (1986) standards (minimum I-CVI of 0.78 by 6 to 10 experts) and a S-CVI/Ave of 0.90 or higher.

Modified kappa statistics

Although scientists often use the CVI to assess content validity, scholars have criticized it for not considering the inflated values that can occur due to the possibility of accidental agreement. Unlike the CVI, the Kappa coefficient eliminates any random chance agreement and improves knowledge of content validity (Zamanzadeh *et al.*, 2014). Therefore, Kappa

statistics may be a significant supplement to – or even a substitute for – the CVI (Wynd, Schmidt, & Schaefer, 2003). The modified Kappa statistic (k^*) adjusts each I-CVI for chance agreement and provides information about expert agreement to express that an item is relevant or clear beyond chance (Polit *et al.*, 2007). The calculation of the modified Kappa statistic requires the calculation of the probability of chance agreement for each scale item, which is expressed in the formula $P_c = [N!/(A!(N - A)!)] \times 0.5^N$, where N is the overall number of experts on a panel and A represents the number of experts who concur that the item is relevant. After that, the Kappa statistic is computed using formula $K = (I-CVI - P_c)/(1 - P_c)$. Values are classified as outstanding, good or fair if they are greater than 0.74, between 0.60 and 0.74 and between 0.40 and 0.59, respectively, in the Kappa evaluation standards (Cicchetti & Sparrow, 1981).

Results of content validity assessment

The experts evaluated 18 items in relation to two attributes (clarity and relevancy). Thus, I calculated two S-CVI/Ave and 36 I-CVI indices, as well as k^* . S-CVI/Ave obtained the value of 0.95 and 0.94 for the clarity and the relevancy rating, respectively and reached the acceptable minimum of 0.90. All 18 items showed excellent validity regarding relevancy (I-CVI ranged from 0.78 to 1 and k^* ranged from 0.76 to 1). Of the 18 items, 17 showed excellent validity in terms of clarity (I-CVI ranged from 0.78 to 1 and k^* ranged from 0.78 to 1). One item (scale item no 6, “I stimulate our team”) had fair validity in terms of clarity (I-CVI = 0.44 and k^* = 0.44) and indicated the need for revision. Table 1 presents the results of the content validity assessment.

The experts commented on all scale items and offered recommendations. Therefore, I conducted a qualitative analysis of the results of the interviews based on feedback and expert comments. These analyses confirmed the results of the quantitative analyses, and therefore the need to correct the scale item no. 6 (“I stimulate our team”) Moreover, they indicated the need to correct scale item no. 18 (“I assess the atmosphere in our team”), which obtained the minimum acceptable value (I-CVI = 0.78) in the quantitative assessment. To revise and complete the scale items, I consulted the experts’ comments, which led to a slight reformulation of two items (nos. 6 “I stimulate our team to act” and 18 “I intuitively sense the atmosphere in our team” to fit them better into the Polish context. Next, I consulted the reformulated scale item with two experts who commented on their clarity and relevance.

Noteworthy, based on the interview findings, I preliminary verified the structure of the TBB construct. I noted experts’ doubts concerning the assignment of several scale items to their corresponding sub-scales. This concerned the following scale items: 9 (“I try to entertain my teammates”), 13 (“I approach my teammates in a personal way”), 15 (“I involve all my teammates in what we do”), 16 (“I respond to my fellow team members’ need”) and 18 (“I assess the atmosphere in our team”). Moreover, the experts commented on the additional statement proposals, which were too heavily profiled in the direction of the team leader. Such profiling is not advisory, because the scale is dedicated to measuring behaviors that can be undertaken by any team member. Consequently, I assumed that the scale would not be expanded to include new items.

Stage 3: psychometric properties of the polish version of the TBB scale: construct validity assessment

Data collection

To assess and validate the reliability and validity of the Polish version of the TBB scale, I conducted a questionnaire survey in Poland in 2022. I decided to use more than one data collection technique in a single survey study comprising the same questionnaire at the

Scale item no.	Item no. in reference*	Clarity rating			Relevance rating			Evaluation		
		N	A	I-CVI	K _{p_c}	K	p _c			
1	9	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
2	8	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
3	7	9	9	1.00	0.001953	1.00	0.89	0.017578	0.89	excellent validity
4	10	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
5	12	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
6	11	9	4	0.44	0.000016	0.44	0.89	0.017578	0.89	excellent validity
7	2	9	9	1.00	0.001953	1.00	0.89	0.017578	0.89	excellent validity
8	5	9	9	1.00	0.001953	1.00	0.89	0.017578	0.89	excellent validity
9	3	9	9	1.00	0.001953	1.00	0.89	0.017578	0.89	excellent validity
10	6	9	9	1.00	0.001953	1.00	0.78	0.070313	0.76	excellent validity
11	1	9	9	1.00	0.001953	1.00	0.89	0.017578	0.89	excellent validity
12	4	9	9	1.00	0.001953	1.00	0.89	0.017578	0.89	excellent validity
13	16	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
14	14	9	8	0.89	0.001953	0.89	1.00	0.001953	1.00	excellent validity
15	18	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
16	15	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
17	13	9	9	1.00	0.001953	1.00	1.00	0.001953	1.00	excellent validity
18	17	9	7	0.78	0.000977	0.78	1.00	0.001953	1.00	excellent validity
		S-CVI/Ave		0.95			S-CVI/Ave		0.94	

Note(s): *Fortuin *et al.* (2021, p. 606)
N number of experts, A experts in agreement, I-CVI item-level content validity index, p_c probability of chance agreement = $[N/(A(N-A))] \times 0.5^N$, K modified Kappa = (I-CVI-p_c)/(1-p_c), S-CVI/Ave scale-level content validity index based on the average method
Source(s): Author's own elaboration

Table 1.
Assessment of content validity

response collection stage (one sample, different techniques) – the optimal research technique for conducting an enterprise research project. I combined computer-assisted telephone interviews (CATI) with online surveys (computer-assisted web interviews, CAWI). This mixed-mode procedure, especially the presence of the CATI technique, provides a number of methodological, organizational and financial advantages that the other available quantitative research techniques lack. Both the CATI and CAWI techniques use a standardized questionnaire as a research tool (i.e. closed questions with a predetermined order and consistent language). The study sample was random. The sampling frame was the Bisnode formerly Hoppenstedt & Bonnier (HBI) database, which contains all the necessary information on business entities operating throughout Poland (including company contact details, Polish Classification of Activities (PKD), employment information, financial data, information on exports and imports). In this context, the research sample was different from the one used for the original scale. The original sample consisted of members of sports teams (74.7%), work teams (17%) and music groups (5.1%). Meanwhile, the Polish sample included members of (work) teams from the business community (100%). This was my deliberate choice, I intended the validated scale to be used mainly in the business environment. A company had to fulfill the following requirements to be included in the sample: having the main seat of the company in Poland and an employment size greater than 50 employees. In the case of the CATI survey, the randomization algorithm built into the software for telephone surveys ensured that each of the records taken from the database of enterprises constituting the so-called gross sample had an identical probability of inclusion in the sample. The response rate was 0.25. Overall, 532 team members participated in the survey.

Sample characteristics

Analyzing the chosen respondents, 62.6% were women and 37.4% were men. The majority of respondents (67.3%) were from medium-sized firms with 50 to 249 employees, while 32.7% were from large organizations with 250 or more employees. The industry demographics were also diverse. Manufacturing (40.2% of respondents) and services (43.6%) were the most common, with trade (16.2%) coming in third. The respondents' tenure as members of the current team ranged from 2 years to 20 years, and the size of the team varied from 2 to 30 people. Almost half (48.3%) of the respondents declared work experience in the team in the range of 6–10 years, 36.7% of them indicated a period from 2 to 5 years, while 15% of the sample declared a period longer than 10 years. The majority of the respondents said that their team consisted of up to 10 people (80.6%).

Measurements

The study utilized structural equation modeling (SEM), a second-generation multivariate method (Chin, 1998). I applied the partial least squares (PLS) method to assess the validity and reliability of the TBB scale, known as a common factor hierarchical second-order model (Becker, Klein, & Wetzels, 2012). The hierarchical nature of the scale favors the PLS method, which avoids the limitations of covariance-based SEM in higher-order construct models, as described in detail by Wetzels, Odekerken-Schröder, and van Oppen (2009). Constraints on sample size, measurement level, model complexity and identification are only a few examples (Wetzels, Odekerken-Schröder, & van Oppen, 2009). PLS can be of use in confirmatory and exploratory studies (Hair, Ringle, & Sarstedt, 2011; Lowry & Gaskin, 2014). It is also widely accepted as a method for testing theory in the early stages, when the research model has not been tested extensively and the theory is less developed (Chin, 1998; Urbach & Ahlemann, 2010; Hair *et al.*, 2011). Moreover, the study found that assumptions regarding the normality of distributions were not met, as indicated by the Jarque–Bera test (Jarque & Bera, 1980; Bera & Jarque, 1981; Kock, 2021), showing that the data were not normal. This further justified the

use of PLS, as it makes no assumptions regarding normal-distributed input data (Wetzels *et al.*, 2009; Urbach & Ahlemann, 2010; Hair *et al.*, 2011).

I performed data analysis using WarpPLS 7.0 software with factor-based outer model analysis algorithm PLS type CFM3 (Kock, 2021). WarpPLS is a nonlinear analysis software tool that allows “nonlinear analyses where best-fitting nonlinear functions are estimated for each pair of structurally linked variables in path models, and subsequently used (i.e. the nonlinear functions) to estimate path coefficients that take into account the nonlinearity” (Kock, 2015, p. 2). This software provides numerous advantages and options for assessing model parameters and computing latent variable scores. Notably, it offers unique features not found in other SEM tools, being the first software to provide both traditional PLS and factor-based algorithms (Kock, 2015, 2021). Factor-based PLS algorithms incorporate covariance-based SEM techniques’ accuracy with the non-parametric properties of conventional PLS algorithms, all while operating under common factor model assumptions (Kock, 2021). Therefore, it helps estimate true composites and factors using factor-based PLS methods that fully account for measurement error. Furthermore, WarpPLS includes a broad range of quality metrics and model fit indicators that are compatible with composite and factor-based SEM (Kock, 2021). The following section will present the results of the measurement model’s reliability and validity in terms of indicator loadings, Cronbach’s alpha, composite reliability, convergent validity and discriminant validity, along with an expanded set of indicators of model fit and quality.

Results of construct validity assessment

According to the standards suggested by Hair, Hult, Ringle and Sarstedt (2017), I evaluated internal consistency reliability, convergent validity and discriminant validity as part of the reflective measurement model evaluation. Internal consistency reliability was examined using Cronbach’s alpha and composite reliability coefficients. For exploratory research, acceptable composite reliability and Cronbach’s alpha should be $\alpha > 0.60$ (Nunnally & Bernstein, 1994; Hair *et al.*, 2017; Kock, 2021). I assessed convergent validity using factor loadings. Scholars recommend two criteria for a measurement model to have acceptable convergent validity, i.e. P values related to the loadings must be equal to or less than 0.05 and the loadings must be equal to or larger than 0.5 (Kock, 2021). Researchers should eliminate indicators with loadings between 0.40 and 0.70 from the scale only if doing so results in a composite reliability increase beyond the specified threshold value (Hair *et al.*, 2011). Occasionally, researchers keep weaker indications because of their contribution to content validity. However, as indicated by Hair *et al.* (2011), scholars should always remove indicators with very low loadings of 0.40 or less from reflective scales. Moreover, to determine convergent validity, researchers must look at the average variance extracted (AVE). According to Fornell and Larcker (1981), an AVE value of 0.50 or greater suggests that the latent variables (constructs) explain more than half of the variability in the indicators. For discriminant validity to exist, the square root of the AVE of each construct must be bigger than other correlations involving that construct (Fornell & Larcker, 1981). Another option to evaluate discriminant validity is the heterotrait–monotrait ratio (HTMT), which must be less than 1.00 (Henseler, Ringle, & Sarstedt, 2015). However, scholars propose the HTMT ratio for assessing discriminant validity in composite-based SEM using classical PLS algorithms, as opposed to factor-based SEM using modern factor estimation algorithms (which I used in this study) (Kock, 2021). Therefore, in this study, the assessment of discriminant validity based on the classical correlations between latent variables and the square roots of the AVE seemed to be sufficient.

The initial model consisting of 18 items had insufficient quality, as highlighted by unacceptable model fit measures and low loadings for items 9 (“I try to entertain my

teammates”), 13 (“I approach my teammates in a personal way”), 15 (“I involve all my teammates in what we do”) and 16 (“I respond to my fellow team members’ needs”) on their corresponding dimensions, reaching the value of <0.4 . As the original model had a weak statistical foundation, I decided to reconfigure the model by eliminating weak test items. The revised model, now with 14 items, demonstrated a satisfactory fit. Table 2 displays the complete results for internal consistency reliability, convergent validity and discriminant validity of the 14-item, three-factor model.

I also used a number of additional indicators to assess model fit (Kock, 2021), including standardized root mean square residual (SRMR), standardized mean absolute residual

1. Internal consistency reliability CR and CA				
Sub-scales	EB	MEB	UB	
CR	0.920	0.902	0.809	
CA	0.920	0.897	0.729	
2. Convergent validity - combined loadings (*p-value <0.00) and AVE				
Scale item no.	Item no. in reference**	EB	MEB	UB
1	9	0.752*		
2	8	0.827*		
3	7	0.836*		
4	10	0.840*		
5	12	0.790*		
6	11	0.822*		
7	2		0.870*	
8	5		0.653*	
10	6		0.842*	
11	1		0.855*	
12	4		0.789*	
14	14			0.911*
17	13			0.913*
18	17			0.405*
AVE		0.659	0.649	0.610
3. Discriminant validity - correlation of latent variables with square root of AVEs				
Sub-scales	EB	MEB	UB	
EB	0.812	-0.037	0.046	
MEB	-0.037	0.806	-0.054	
UB	0.046	-0.054	0.781	

4. Model fit indices

SRMR = 0.104, acceptable if ≤ 0.1 ; SMAR = 0.075, acceptable if ≤ 0.1 ; SChS = 1.334, $P < 0.001$; STDCR = 0.945, acceptable if ≥ 0.7 , ideally = 1; STDSR = 0.783, acceptable if ≥ 0.7 , ideally = 1; AFVIF = 1.004, acceptable if ≤ 5 , ideally ≤ 3.3

Note(s): **Fortuin *et al.* (2021), p. 606

EB energizing behaviors, MEB mood-enhancing behaviors, UB uniting behaviors, CR composite reliability, CA Cronbach’s alpha, AVE average variance extracted, SRMR standardized root mean square residual, SMAR standardized mean absolute residual, SChS standardized chi-squared, STDCR standardized threshold difference count ratio, STDSR standardized threshold difference sum ratio, AFVIF average full collinearity VIF

Source(s): Author’s own elaboration

Table 2.
Results of the
validation study

(SMAR), standardized chi-squared (SChS), standardized threshold difference count ratio (STDSCR) and standardized threshold difference sum ratio (STDSR). I obtained the following satisfactory values for: SRMR = 0.104, acceptable if ≤ 0.1 ; SMAR = 0.075, acceptable if ≤ 0.1 ; SChS = 1.334, $P < 0.001$; STDSCR = 0.945, acceptable if ≥ 0.7 and ideally = 1; STDSR = 0.783, acceptable if ≥ 0.7 and ideally = 1; average full collinearity VIF (AFVIF) = 1.004, acceptable if ≤ 5 and ideally ≤ 3.3 .

Discussion

I aimed to adapt and validate the Polish version of the TBB scale originally developed by Fortuin *et al.* (2021). The modified and reduced 14-item Polish adaptation of the TBB exhibited appropriate psychometric qualities, making it a useful tool for evaluating TBBs in organizations within a Polish sociocultural environment. However, it is crucial to understand the broader reasons for removing four items from the original scale due to their minimal loading on dimensions. Noteworthy, by the interview stage of the qualitative study, experts faced difficulty accurately assigning the removed items to their sub-scales, as mentioned earlier. This suggests potential issues with scale structure, dimension naming accuracy and cultural differences. The resulting scale structure proved to be in line with my expectation that differences in Hofstede's cultural dimensions between Polish and Dutch culture may result in differences in the perception of TBB by Polish versus Dutch experts and respondents. I expected differences especially in the areas of mood-enhancing and uniting behaviors. And it was precisely the items from these areas that received the lowest loadings. Discrepancies may also arise from the research context. My research sample differed from the one used by the authors of the original scale. I surveyed only members of business teams, whereas the original sample also included members of sports and music groups, who together made up the majority. Using business teams as the target group in this study may have certain implications. Although the literature indicates that in the case of organizational work teams, sports teams can serve as a model (Katz, 2001), it is important to consider the similarities and divergences of sports teams and non-sports teams to effectively use sports teams to expand the knowledge of organizational work teams (Wolfe *et al.*, 2005). There are certain similarities between these two team types. Both types of teams work in intensely competitive contexts with well-defined performance measures and both have decision-makers who choose and put into practice tactics to increase competitiveness (Wright, Smart, & McMahan, 1995). In terms of differences, sports teams and other organizational teams clearly differ in how they monitor resources and maintain relationships between members (Mach, Dolan, & Tzafir, 2010). Compared to business teams, sports teams have a higher level of commitment because they typically have definite, difficult and shared objectives, such as winning a championship (Hakanen, Häkkinen, & Soudunsaari, 2015). In sports teams, members realize that the objective is to win and that the way to do so is typically straightforward and obvious (Katz & Koenig, 2001). The situation is different in the workplace, where goals are not well defined and change regularly, and organizations struggle to explain the plan of action to all team members (Katz & Koenig, 2001). Business teams frequently lack clearly defined objectives, which lowers personal commitment (Mach *et al.*, 2010). For example, another difference is that sports teams essentially operate within the two different modes of play and training, as games alternate with training (Katz, 2001). During games, the team is in performance mode, while during training, it is in "learning" mode and covers up mistakes. In contrast, teams at work are frequently under so much pressure to achieve outcomes that they are constantly in performance mode, neglecting the need to balance this mode with learning mode (Katz, 2001). These differences between sports teams and business teams may be the reason for the different results obtained in assessing the individual items of the original scale and its Polish-validated version. Teams operating in a

business environment and their dynamics may differ from other organizational teams, such as sports teams. Sports teams display a different level of team chemistry due to a higher level of commitment to a common, clear goal of winning. Meanwhile, in sports teams, there might be more time for team-building activities during their learning phase, which is not typically present in business teams. This may make TBB “uniting” sub-scale ratings different in samples of respondents representing business and sports teams, as demonstrated in this study. I advise further research in this area.

Conclusions and implications

The adapted version of the TBB scale is a valid and reliable research instrument for conducting empirical research in the area of organizational behavior and, in particular, for determining the importance of individual team-strengthening behaviors in defining team dynamics and team performance. Therefore, the data it offers can significantly aid in understanding the psychological nature of the workplace. The presented research results are an important contribution, as having a valid and reliable instrument available in several languages promotes international cooperation and synergy. It will also help to fill a knowledge gap and build a link between academics, businesspeople and labor experts working in many other organizations to better understand the psychological work environment. My research also provides practical implications. The scale authors posit that “team boosting behaviors are developable and trainable and can be promoted by managers or organizational practices” (Fortuin *et al.*, 2021, p. 615). If future research confirms that TBBs positively impact team dynamics and performance, team leaders might consider (1) adopting the role of a team booster by incorporating behaviors from the TBB scale into daily management, (2) providing training on TBBs to specific team members, or (3) incorporating team boosting skills as a requirement in recruitment processes. With the new version of the validated TBB scale, this could become a reality in yet another country, expanding the share of positive behavior in the functioning of teams, in line with the POS concept.

Limitations and future research

The study displays several advantages, such as the qualitative research with experts preceding the quantitative validation research, using the SEM technique to assess the scale’s psychometric properties and ensure the result’s validity and reliability as well as the use of random sampling, which allows for the reduction of bias. However, it also has some limitations. First, the study’s focus on business teams might limit external validity, cautioning against broad extrapolation to other industries or organizational teams (e.g. top management, improvement, or research teams) with diverse characteristics. I strongly advise further research for more comprehensive insights. Second, this article was limited to assessing the psychometric properties of the Polish version of the TBB scale and provided preliminary evidence for different forms of reliability and validity. I applied a post-hoc model modification by redefining the measurement tool. This method helped find factorial confounding components that may be eliminated from the model. Noteworthy, the experts and practical performers did not further verify the reduced scale after stage 3, which evaluated the psychometric properties of the Polish version of the TBB scale. This limitation sets the stage for further work that would be worthwhile with this research tool.

Moreover, the use of both the validated and the original scale requires great caution from researchers. The TBBs construct is one of the collective team constructs measured using aggregate scores based on each team member’s data, i.e. data at the individual level. To effectively examine team-level constructs, researchers must have a thorough understanding

of data aggregation techniques and key analytic approaches. Multilevel literature can be helpful for this purpose (Bliese, 2000; Chan, 2005).

To offer valuable recommendations for future research and address the fundamental question of whether a construct enables accurate predictions, it is essential to evaluate the instrument's nomological validity (Czakon, 2019). In particular, it would be crucial to check how it performs with other theoretically grounded metrics. Therefore, following this study, research work would have to focus on finding the relationship between the adapted TBB and other theoretically consistent constructs, where higher-order constructs such as teamwork engagement or team job crafting could be included in the modeling. Future research should also focus on establishing additional degrees of criterion validity, reliability and test–retest use. Further replication studies should confirm the scale's usefulness in various organizations and industries, considering a comprehensive evaluation of the tool's psychometric characteristics.

References

- Aycan, Z., & Gelfand, M. J. (2012). Cross-cultural organizational psychology. In S. W. J. Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology*. Oxford University Press. doi: [10.1093/oxfordhb/9780199928286.013.0033](https://doi.org/10.1093/oxfordhb/9780199928286.013.0033).
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measure. *Spine*, *25*(24), 3186–3191. doi: [10.1097/00007632-200012150-00014](https://doi.org/10.1097/00007632-200012150-00014).
- Becker, J. M., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: Guidelines for using reflective-formative type models. *Long Range Planning*, *45*(5-6), 359–394. doi: [10.1016/j.lrp.2012.10.001](https://doi.org/10.1016/j.lrp.2012.10.001).
- Bera, A. K., & Jarque, C. M. (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters*, *7*(4), 313–318. doi: [10.1016/0165-1765\(81\)90035-5](https://doi.org/10.1016/0165-1765(81)90035-5).
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods in Organizations. Foundations, Extensions, and New Directions* (pp. 249–381). San Francisco: Jossey-Bass.
- Cameron, K. S., & Caza, A. (2004). Contributions to the discipline of positive organizational scholarship. *American Behavioral Scientist*, *47*(6), 731–739. doi: [10.1177/0002764203260207](https://doi.org/10.1177/0002764203260207).
- Cameron, K. S., Dutton, J. E., & Quinn, R. E. (2003). *Positive organizational scholarship: Foundations of a new discipline*. San Francisco: Berrett-Koehler.
- Chan, D. (2005). Multilevel research. In F. T. L. Leong, & J. T. Austin (Eds.), *The Psychology Research Handbook* (2nd ed, pp. 401–18). Thousand Oaks, CA: Sage.
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Lawrence Erlbaum Associates.
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*(2), 127–37.
- Collins, A. L., Lawrence, S. A., Troth, A. C., & Jordan, P. J. (2013). Group affective tone: A review and future research directions. *Journal of Organizational Behavior*, *34*(1), 43–62, doi: [10.1002/job.1887](https://doi.org/10.1002/job.1887).
- Country Comparison Tool (2023), Available from: <https://www.hofstede-insights.com/country-comparison-tool>
- Czakon, W. (2019). Walidacja narzędzia pomiarowego w naukach o zarządzaniu. *Przegląd Organizacji*, *4*(951), 3–10. doi: [10.33141/po.2019.04.01](https://doi.org/10.33141/po.2019.04.01).

-
- Davis, L. L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research, 5*(4), 194–197. doi: [10.1016/S0897-1897\(05\)80008-4](https://doi.org/10.1016/S0897-1897(05)80008-4).
- DeVellis, R. F. (2016). *Scale development: Theory and applications* (Fourth Edition). Los Angeles, CA: Sage publications.
- Felps, W., Mitchell, T. R., & Byington, E. (2006). How, when, and why bad apples spoil the barrel: Negative group members and dysfunctional groups. *Research in Organizational Behavior, 27*, 175–222. doi: [10.1016/S0191-3085\(06\)27005-9](https://doi.org/10.1016/S0191-3085(06)27005-9).
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. doi: [10.2307/3151312](https://doi.org/10.2307/3151312).
- Fortuin, D. J., van Mierlo, H., Bakker, A. B., Petrou, P., & Demerouti, E. (2021). Team boosting behaviors: Development and validation of a new concept and scale. *European Journal of Work and Organizational Psychology, 30*(4), 600–618. doi: [10.1080/1359432X.2020.1854226](https://doi.org/10.1080/1359432X.2020.1854226).
- Gibson, C. B. (1999). Do they do what they think they can? Group-Efficacy and group effectiveness across tasks and cultures. *Academy of Management Journal, 42*(2), 138–152. doi: [10.2307/257089](https://doi.org/10.2307/257089).
- Glińska-Noweś, A. (2017). *Pozytywne relacje interpersonalne w zarządzaniu*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika w Toruniu.
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology, 61*(2), 29–145. doi: [10.1027/1901-2276.61.2.29](https://doi.org/10.1027/1901-2276.61.2.29).
- Haffer, R., & Glińska-Noweś, A. (2013). Pozytywny Potencjał Organizacji jako determinanta sukcesu przedsiębiorstwa. Przypadek Polski i Francji. *Zarządzanie i Finanse, 11*(4), 91–1100.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139–152. doi: [10.2753/MTP1069-6679190202](https://doi.org/10.2753/MTP1069-6679190202).
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles, CA: Sage Publications.
- Hakanen, M., Häkkinen, M., & Soudunsaari, A. (2015). Trust in building high-performing teams: Conceptual approach. *Electronic Journal of Business Ethics and Organization Studies, 20*(2), 43–53.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum.
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238–247. doi: [10.1037/1040-3590.7.3.238](https://doi.org/10.1037/1040-3590.7.3.238).
- Heggstad, E. D., Scheaf, D. J., Banks, G. C., Hausfeld, M. M., Tomidandel, S., & Williams, E. B. (2019). Scale adaptation in organizational science research: A review and best-practice recommendation. *Journal of Management, 45*(6), 2596–2627. doi: [10.1177/0149206319850280](https://doi.org/10.1177/0149206319850280).
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. doi: [10.1007/s11747-014-0403-8](https://doi.org/10.1007/s11747-014-0403-8).
- Hofstede, G. (1980). Motivation, leadership and organization: Do American theories apply abroad?. *Organizational Dynamics, 9*(1), 42–63. doi: [10.1016/0090-2616\(80\)90013-3](https://doi.org/10.1016/0090-2616(80)90013-3).
- Hofstede, G., Hofstede, G. J., & Minkov, M. (1991). *Software of the mind: Cultures and organizations*. London: McGraw-Hill.
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255–259. doi: [10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5).

-
- Katz, N. (2001). Sports teams as a model for workplace teams: Lessons and liabilities. *Academy of Management Perspectives*, 15(3), 56–67. doi: [10.5465/ame.2001.5229533](https://doi.org/10.5465/ame.2001.5229533).
- Katz, N., & Koenig, G. (2001). Sports teams as a model for workplace teams: Lessons and liabilities [and executive commentary]. *The Academy of Management Executive*, 15(3), 1993–2005, 56–69.
- Kirkman, B. L., Gibson, C. B., & Shapiro, D. L. (2001). 'Exporting' teams: Enhancing the implementation and effectiveness of work teams in global affiliates. *Organizational Dynamics*, 30(1), 12–29. doi: [10.1016/s0090-2616\(01\)00038-9](https://doi.org/10.1016/s0090-2616(01)00038-9).
- Kock, N. (2015). A note on how to conduct a factor-based PLS-SEM analysis. *International Journal of e-Collaboration*, 11(3), 1–9. doi: [10.4018/ijec.2015070101](https://doi.org/10.4018/ijec.2015070101).
- Kock, N. (2021). *WarpPLS user manual: Version 7.0*. Laredo, TX: ScriptWarp Systems.
- Lonner, W., & Malpass, R. S. (1994). When psychology and culture meet: An introduction to cross-cultural psychology. In W. J. Lonner, & R. S. Malpass (Eds.), *Psychology and culture* (pp. 1–12). Boston: Allyn & Bacon.
- Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *Iee Transactions on Professional Communication*, 57(2), 123–146. doi: [10.1109/TPC.2014.2312452](https://doi.org/10.1109/TPC.2014.2312452).
- Luthans, F. (2002). The need for and meaning of positive organizational behavior. *Journal of Organizational Behavior*, 23(6), 695–706. doi: [10.1002/job.165](https://doi.org/10.1002/job.165).
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. doi: [10.1097/00006199-198611000-00017](https://doi.org/10.1097/00006199-198611000-00017).
- Mach, M., Dolan, S., & Tzafirir, S. (2010). The differential effect of team members' trust on team performance: The mediation role of team cohesion. *Journal of Occupational and Organizational Psychology*, 83(3), 771–794. doi: [10.1348/096317909X473903](https://doi.org/10.1348/096317909X473903).
- Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997–2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, 34(3), 410–476. doi: [10.1177/0149206308316061](https://doi.org/10.1177/0149206308316061).
- Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and integration of team composition models: Moving toward a dynamic and temporal framework. *Journal of Management*, 40(1), 130–160. doi: [10.1177/0149206313503014](https://doi.org/10.1177/0149206313503014).
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing Health*, 29(5), 489–497. doi: [10.1002/nur.20147](https://doi.org/10.1002/nur.20147).
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing Health*, 30(4), 459–467. doi: [10.1002/nur.20199](https://doi.org/10.1002/nur.20199).
- Rubio, D., Berg-Weger, M., Tebb, S. S., Lee, S. E., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(3), 94–104. doi: [10.1093/swr/27.2.94](https://doi.org/10.1093/swr/27.2.94).
- Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*, 54(4), 599–622. doi: [10.1002/hrm.21628](https://doi.org/10.1002/hrm.21628).
- Schneider, S. C., & Barsoux, J. L. (2003). *Managing across cultures*. New York: Pearson Education.
- Shrotryia, V. K., & Dhanda, U. (2019). *Content validity of assessment instrument for employee engagement* (pp. 1–7). Los Angeles: SAGE Open. doi: [10.1177/2158244018821751](https://doi.org/10.1177/2158244018821751).
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268–274. doi: [10.1111/j.1365-2753.2010.01434.x](https://doi.org/10.1111/j.1365-2753.2010.01434.x).

-
- Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using partial least squares. *Journal of Information Technology Theory and Application*, *11*(2), 5–40.
- Walter, F., & Bruch, H. (2008). The positive group affect spiral: A dynamic model of the emergence of positive affective similarity in work groups. *Journal of Organizational Behavior*, *29*(2), 239–261. doi: [10.1002/job.505](https://doi.org/10.1002/job.505).
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2005). *Measurement in nursing and health research*. New York: Springer Publishing Company.
- Wetzels, M., Odekerken-Schröder, G., & van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *Management Information Systems Quarterly*, *33*(1), 177–195. doi: [10.2307/20650284](https://doi.org/10.2307/20650284).
- Wolfe, R. A., Weick, K. E., Usher, J. M., Terborg, J. R., Poppo, L., Murrell, A. J., . . . Jourdan, J. S. (2005). Sport and organizational studies: Exploring synergy. *Journal of Management Inquiry*, *14*(2), 182–210. doi: [10.1177/1056492605275245](https://doi.org/10.1177/1056492605275245).
- Wright, T. A. (2003). Positive organizational behavior: An idea whose time has truly come. *Journal of Organizational Behavior*, *24*(4), 437–442. doi: [10.1002/job.197](https://doi.org/10.1002/job.197).
- Wright, P. M., Smart, D. L., & McMahan, G. C. (1995). Matches between human resources and strategy among NCAA basketball teams. *Academy of Management Journal*, *38*(4), 1052–1074. doi: [10.2307/256620](https://doi.org/10.2307/256620).
- Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, *25*(5), 508–518. doi: [10.1177/0193945903252998](https://doi.org/10.1177/0193945903252998).
- Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H. A., Nikanfar, A., & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today*, *1*(3), 163–171. doi: [10.18502/npt.v7i1.2295](https://doi.org/10.18502/npt.v7i1.2295).

Corresponding author

Joanna Haffer can be contacted at: joanna.haffer@torun.merito.pl