Journal of Interprofessional Care

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ijic20

Validation of a French-language version of TeamSTEPPS® T-TPQ and T-TAQ questionnaires

Anh Nguyet Diep, Méryl Paquay, Jean-Christophe Servotte, Nadia Dardenne, Laure Istas, Pauline Van Ngoc, Mathilde Charlier, Alexandre Ghuysen & Anne-Françoise Donneau

To cite this article: Anh Nguyet Diep, Méryl Paquay, Jean-Christophe Servotte, Nadia Dardenne, Laure Istas, Pauline Van Ngoc, Mathilde Charlier, Alexandre Ghuysen & Anne-Françoise Donneau (2022) Validation of a French-language version of TeamSTEPPS® T-TPQ and T-TAQ questionnaires, Journal of Interprofessional Care, 36:4, 607-616, DOI: 10.1080/13561820.2021.1902293

To link to this article: https://doi.org/10.1080/13561820.2021.1902293

JOURNAL OF INTERPROFESSIONAL CARE
2022, VOL. 36, NO. 4, 607–616
https://doi.org/10.1080/13561820.2021.1902293

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Validation of a French-language version of TeamSTEPPS® T-TPQ and T-TAQ questionnaires

Anh Nguyet Diep [a,b1], Méryl Paquay[c1], Jean-Christophe Servotte[d], Nadia Dardenne[a], Laure Istas[e], Pauline Van Ngoc[f], Mathilde Charlier[a], Alexandre Ghuysen[a,c2], and Anne-Françoise Donneau[a2]

[a]Public Health Sciences Department, University of Liège, Liège, Belgium; [b]Information Technology Department, Can Tho University, Can Tho, Vietnam; [c]Medical Simulation Center of Liège, University of Liège, Public Health Sciences Department, Liège, Belgium; [d]Cellule Recherche & Développement, Nursing Department, Centre Namur Simulation, Liège, Belgium; [e]Plateforme Pour l'Amélioration Continue de la Qualité des Soins et de la Sécurité des Patients, Brussels, Belgium; [f]Unité de Recherche Soins Primaires et Santé Département de Médecine Générale, University of Liège, Liège, Belgium

ABSTRACT

Teamwork training and evaluation are essential to enhance safety and quality of care. The lack of the psychometric testing of the TeamSTEPPS® Teamwork Attitudes Questionnaire (T-TAQ) and Teamwork Perceptions Questionnaire (T-TPQ) across different language and cultural settings has questioned their widespread use because such attitudes and perceptions are highly subjective and context-bound. The present study aims to translate the T-TAQ and T-TPQ into the French language and validate the psychometric properties of the two questionnaires in a public health context. A forward-backward translation process, panel reviewing, and pilot testing in two rounds were followed to develop the French versions. Confirmatory factor analysis (CFA) and Cronbach's alpha were used to examine the factor structure and internal consistency, whereas two-way mixed Intraclass Correlation Coefficient (ICC) was performed to assess test-retest reliability. A total of 235 healthcare professionals in the French-speaking community of Belgium completed the T-TAQ and T-TPQ. After two to four weeks, 102 participants took part in the second round. Despite good fit indices as revealed by the CFA and Cronbach's alpha from 0.53 to 0.75 for the five dimensions of the T-TAQ and 0.76 to 0.79 for the T-TPQ, the squared correlations among the constructs were higher than the average variance extracted. Two-way mixed ICCs indicated fair to good test-retest reliability for all the five constructs of the two questionnaires, except the leadership scale of the T-TAQ. The French-language versions of the T-TAQ and T-TPQ were semantically equivalent and culturally relevant with adequate test-retest reliability as compared to the English versions. These two instruments might be used to capture the overall attitude toward teamwork and perceptions of team skills and behaviors. Yet, further research is advisable to refine the scales to establish the discriminant validity of the different dimensions and discriminative power of the instruments.

ARTICLE HISTORY
Received 16 November 2020
Revised 28 February 2021
Accepted 05 March 2021

KEYWORDS
Team training; patient safety; interdisciplinary; surveys; teamwork

Introduction

In its report "To err is human", the Institute of Medicine (IOM) estimated that between 44,000 and 98,000 patients died each year due to medical errors (Kohn et al., 2000). These figures drew attention to the quality of care and patient safety, laying the foundations for movement such as the Institute for Healthcare Improvement's 100,000 Lives Campaign (Peberdy et al., 2003). A key priority of this campaign lies in the implementation of Rapid Response Systems (RRS) including Rapid Response Teams (RRT). This strategy aims at preventing medical errors through the reduction of patients' unmet needs preceding cardiorespiratory arrest (Baker et al., 2005). Accordingly, the RRS and RRT would help improve early symptoms detection and overcome challenges impeding rapid and effective response (Winters et al., 2013). Given the impact on patient safety, RRT implementation has been highly recommended by leading authorities such as the Joint Commission (Silva et al., 2016). Teamwork training is however a prerequisite for an RRT to be optimal. Indeed, working in teams in the healthcare mobilizes several skills, behaviors, cognitive frames and attitudes that mitigate the efficiency and safety of care (Flin et al., 2008).

For this purpose, in 2007, the US Agency for Healthcare Research and Quality (AHRQ) and Department of Defense (DoD) developed a teamwork training entitled TeamSTEPPS® (Team Strategies and Tools to Enhance Performance and Patient Safety) (Baker et al., 2007). Based on more than 30 years of team training and teamwork, it aimed to improve four teamwork skills: leadership, situation monitoring, mutual support, and communication. Studies conducted to assess the effectiveness of teamwork training revealed an improvement in patient safety culture and quality of care (Capella et al., 2010; Mayer et al., 2011; Thomas & Galla, 2013). However, they also pointed to the current lack of validated standardized tools available to assess teamwork quality and effectiveness (Baker et al., 2010; Manser, 2009).

CONTACT Anh Nguyet Diep anhnguyet.diep@uliege.be Public Health Sciences Department, University of Liège, Liège, Belgium.
[1]Co-first authors.
[2]Co-last authors.
Supplemental data for this article can be accessed on the publisher's website.

As a result, Baker et al. (2010) developed the TeamSTEPPS® Teamwork Attitudes Questionnaire (T-TAQ) to assess individual attitudes related to the role of teamwork. This questionnaire includes 30 items, around five dimensions of teamwork: team structure, leadership, situation monitoring, mutual support, and communication. A second questionnaire was developed to assess the perception of team skills and behaviors (TeamSTEPPS® Teamwork Perceptions Questionnaire – T-TPQ). This questionnaire involves five dimensions of teamwork: structure, leadership, situation monitoring, mutual support and communication, representing the basic skills for teamwork (Keebler et al., 2014) that would most affect team performance (Salas et al., 2005).

While these two questionnaires were validated in English, they have already been translated into various languages such as Norwegian (Ballangrud et al., 2017), Portuguese (Brandão et al., 2016) and Korean (Hwang & Ahn, 2015). However, to date, no French-language version has been validated. As French is the fifth most widely spoken language in the world, it is justified to carry out the French-language translation and validation of the T-TAQ and T-TPQ for wider usage among the French-speaking scientific community.

The present study aims to (1) translate the T-TAQ and T-TPQ into the French language using a forward-backward translation process and (2) validate the psychometric properties of the two questionnaires in terms of content validity, factor structure, discriminant validity, and test-retest reliability.

Methodology

French-language version of the T-TPQ and the T-TAQ: the forward-backward translation process

To realize the final French versions of the T-TPQ and T-TAQ, we followed the forward-backward translation process as recommended by Beaton et al. (2000) and Sousa and Rojjanasrirat (2011). This included the steps further elaborated as follows.

Forward translation

The original versions of the T-TPQ and the T-TAQ were independently translated to French by a professional translator and one bilingual translator fluent in both French and English. They had French as their native language and were familiar with the field of health care.

Panel reviewing

A French-speaking review committee consisted of one occupational psychologist, two clinical psychologists, one doctor in biomedical and pharmaceutical sciences and one public health nurse, working in the healthcare fields, met to reach a consensus on the translated items. These people were selected on the basis of their knowledge and experience. The committee worked together to discuss and improve the clarity, coherence, and cultural relevance of the preliminary-translated versions.

Back-translation

A professional bilingual translator with English as mother tongue back-translated the French version approved by the review committee into English. This back-translation was done without the help of the original English version.

Comparison of the back-translated version and the original English version: content validity

The back-translated and the original versions were submitted to the above-mentioned review committee. They compared the two versions regarding the grammatical structure of the items, the similarity of meaning, and the relevance. The participants could express their opinions regarding the relevance of the items through several meetings that were organized. A consensus was obtained if all group members gave a positive answer to all the three questions. In doing so, content validity could be established (Lynn, 1986).

Pilot testing

A pilot test was organized to assess the clarity of each item of the T-TPQ and the T-TAQ. The initial pilot test was performed on a sample of 30 participants who were working as nurses and physicians at the University Hospital of Liège, with a mean age of 29.7 ± 9.96 years. Participants were asked to rate the clarity of each item through a dichotomous scale (0 = "not clear", 1 = "clear"). For items that did not reach the 80% of "clear" answers, a second pilot test with 15 participants drawn from the same sample was conducted. This 80% figure was used to ensure the linguistic and semantic equivalence of the tool and thus, making it possible to further improve the structure so as to be easily understood by the target population before psychometric property validation (Sousa & Rojjanasrirat, 2011).

Testing the psychometric properties of the French version of T-TPQ and the T-TAQ

In order to examine the psychometric properties of the two questionnaires, a cross-sectional design with convenience sampling was employed. The study design was reviewed and approved by the university ethics committee (reference number 2018/340 dated December 4, 2018).

Sample and procedure

Participants (N = 255) were recruited among healthcare professionals from different departments from four hospitals in the French-speaking region of Belgium (Center Hospitalier Universitaire de Liège – CHU Liège), Center Hospitalier Régional de Liège – CHR Liège, Center Hospitalier Régional de Sambre et Meuse – CHR Sambre et Meuse, and Hôpital Erasme – Cliniques Universitaires de Bruxelles – CUB Erasme). In general, the participants were registered nurses and residents working in the Intensive Care Service,

Emergency Department, and non-acute hospital nursing unit; and nurses following a master program in Public Health at the University of Liège. None of them were administrative staff. In order to minimize issues of common method bias, we followed the recommendations from MacKenzie and Podsakoff (2012). More specifically, a lot of effort was invested to enhance the participants' motivation and ability to answer the questionnaire by stressing the significance of their answers and emphasizing that there were no right or wrong answers. Anonymity was ensured and the voluntary nature of the survey participation was clearly communicated to the participants so that no threats related to participation was entailed. Four participants refused to respond and six did not speak French. The data cleaning process resulted to 10 cases with missing data and were removed from the analysis, yielding a sample size of 235 for Phase 1. After an interval of two to four weeks, a second invitation was sent to ask the participants for their responses. One hundred and two of them responded (41.6%), whose scores from both phases were paired and used to evaluate the test-retest reliability.

Data recording

Every participant completed a general information section (e.g., gender, diploma). The participants were then invited to give their responses regarding the attitude and perceptions of teamwork, measured by the T-TAQ and T-TPQ.

The T-TAQ (Baker et al., 2010) consists of 30 items capturing 5 dimensions of teamwork attitude, namely team structure, leadership, situation monitoring, mutual support, and communication. All the items were anchored on a five-point Likert scale from (1) strongly disagree to (5) strongly agree. Four items, i.e. i20, i21, i24 of the mutual support and i30 of the communication scale, were reversed coded.

The T-TPQ validated by Keebler et al. (2014) comprises 35 items measuring 5 dimensions of perceptions regarding team skills and behaviors: team structure, leadership, situation monitoring, mutual support, and communication. Each dimension was measured on five-point Likert scale from (1) strongly disagree to (5) strongly agree.

According to the TeamSTEPPS® manual, either a total sum score or an average score is calculated for each dimension. To be comparable with other validation studies, the average approach was adopted in the present study. For both scales, higher scores indicate more positive responses whereas lower scores present an unfavorable attitude and perception toward teamwork (Baker et al., 2008).

Statistical analysis

Descriptive statistics were performed to describe the sample characteristics. Shapiro-Wilk tests were conducted to examine the normality of the scores. On this basis, either the mean and standard deviations (SD) or the median and interquartile ranges (1st and 3rd quartiles) were reported.

Confirmatory factor analysis (CFA) was run to examine the factor structure of the T-TAQ and T-TPQ. Chi-square test and Root Mean Square Error of Approximation (RMSEA<0.06)), Comparative Fit Index (CFI>0.90), Tucker-Lewis Index (TLI>0.90), and Standardized Root Mean Square Residual (SRMR<0.08) were employed as goodness of fit indices (Hu & Bentler, 1999). A non-significant chi-square test was evidence of good model fit, but this test was highly sensitive to sample size. Therefore, a significant result did not effectively suggest poor model fit (Harlow, 2014). Alternatively, the normed chi-square, i.e., the ratio between the actual chi-square value and degree of freedom, was recommended with values from one to three being considered acceptable. Royston tests and Q-Q plots were employed to examine the multivariate normality of the variables. If the assumptions were violated, robust maximum likelihood estimation with robust (Huber-White) standard errors and a scaled test statistic were performed for the CFA. Post-hoc modifications of the model were carried out in case of inadequate factor loadings or fit indices to examine where the problem was inherent. Next, Cronbach's alpha (α) was calculated for each established construct to assess the internal consistency.

The average variance extracted (AVEs) of each construct was compared with the squared correlations among them to examine the discriminant validity or construct independence, AVEs should be above 0.5 and larger than the squared correlations as evidence of discriminant validity. Test-retest reliability of each established factor of the two questionnaires was then explored by Intraclass correlation coefficients (ICCs) (two-way mixed, absolute agreement). Standard error of measurement (SEM) as measure of absolute reliability and minimum difference were calculated to provide an overview about the variations between the observed scores and the true scores. The higher the minimum difference, the larger the variations compared with the true scores (Hopkins, 2000), and the higher the minimum difference, the more substantial change that should occur to go beyond the measurement error (Weir, 2005). Floor and ceiling effects, which demonstrate the discriminative power of the two instruments, were examined. Finally, Kruskal-Wallis tests with Dunn's post-hoc tests were conducted to examine the differences in the scoring across the dimensions of the T-TAQ and T-TPQ between males and females as well as among holders of different educational degrees. The analyses were conducted in R (R Core Team, 2013).

Results

The translation of the French version of the T-TPQ and the T-TAQ

Content validity

The review committee compared the back-translated version and the original English version of the T-TPQ and the T-TAQ based on three criteria, namely semantic equivalence, relevance, and ambiguity by means of three guided questions. This process resulted in 16 differences between original and the back-translated version for the T-TAQ and 14 differences for the T-TPQ. After the panel meeting, the final modifications were realized. The result was presented in Table 1.S in the Supplementary material.

Pilot testing

In the first round, out of the 30 items of the T-TAQ, 23 items were considered as being clear in meaning. For the T-TPQ, 32 out of 35 items were considered as being clear. Thus, modifications were needed before the second testing was conducted and were found in Table 2.S in the Supplementary material.

In the second round, the 10 modified items were rated by 15 participants, all of which received satisfactory consensus on the clarity. This resulted in the definitive French versions of the T-TAQ and T-TPQ, which can be found in the 3.S and 4.S in the Supplementary material, respectively.

The psychometric properties of the TPQ and the T-TAQ

Participant characteristics

In Phase 1, the participants were from different units of CHU Liège (60.8%) and CUB Erasme (33.2%). More than three-quarters of them were female. Due to the small number of participants within each educational level, we collapsed secondary and professional bachelor into pre-graduate group; and master and doctoral degrees into post-graduate group while bachelor's degree holders group remained. The latent scores were obtained for five constructs of the T-TAQ and T-TPQ. Furthermore, one participant reporting with missing data on diploma was removed from the analysis. As to education attainment, 72.8% participants obtained a bachelor degree, followed by post-graduate degree (19.1%) and pre-graduate degree (7.7%) holders. In Phase 2, the tendency was similar, with the majority of the participants was female (76.5%), holding a bachelor degree (68.6%), and coming from CHU Liège (49.0%) and CUB Erasme (49.0%). The socio-demographics can be found in Table 1.

Descriptive statistics of item ratings

All items of the T-TAQ and T-TPQ displayed a skewed distribution. The results in Tables 5.S in the Supplementary material presented the median and IQR of the items in the first round.

Table 1. Respondents' socio-demographics (N = 235).

		Phase 1	Phase 2
Characteristic	Categories	Number (%)	Number (%)
Gender	Male	53 (22.6)	24 (23.5)
	Female	182 (77.5)	78 (76.5)
Diploma	*Pre-graduate*	*18 (7.7)*	*11 (10.8)*
	Secondary diploma	10 (4.3)	6 (5.9)
	Professional bachelor (Brevet)*	8 (3.4)	5 (4.9)
	Bachelor degree	*171 (72.8)*	*70 (68.6)*
	Post-graduate	*45 (19.1)*	*21 (20.6)*
	Master	40 (17.0)	20 (19.6)
	Doctoral degree	5 (2.1)	1 (1.0)
	Missing	*1 (0.4)*	*0 (0.00)*
Institution	CHU Liège	143 (60.8)	50 (49.0)
	CHR Liège	10 (4.3)	2 (2.0)
	CHR Sambre et Meuse	4 (1.7)	0 (0.00)
	CUB Erasme	78 (33.2)	50 (49.0)

Notes. * Compare to a Bachelor, the Brevet focuses more on practice and clinical care. The Brevet does not allow access to all nursing specializations, e.g., the degree holders could not progress to emergency and intensive care specialization.

Confirmatory factor analysis (CFA)

Thirty items constituting the T-TAQ and the 35 items of the T-TPQ did not display multivariate normality with R = 2227.60, $p < .0001$ and R = 2213.17, $p < .0001$, respectively for the Royston tests results. Q-Q plots also confirmed this deviation from normality. As the multivariate normality assumption was violated, CFA was conducted with robust methods for both the T-TAQ and T-TPQ.

Confirmatory Factor analysis (CFA) for the T-TAQ. The factor structure of the T-TAQ was examined in light of the robust fit indices. The result showed that the model displayed in adequate fit with $\chi2(395) = 653.473$, $p < .001$ (normed chi-square = 1.65, falling within the recommended range from one to three), CFI = 0.796, TLI = 0.776, RMSEA = 0.050 and SRMR = 0.069. The factor loadings revealed that i20, i21, and i30 had loadings lower than 0.320. Thus, we attempted to rerun the CFA by removing one item at a time, one of two the pairs at a time, and all three variables together to examine if all the remaining items could result in better fit and factor loadings. This entailed six post-hoc examinations of the model. This iterative process suggested removing all the three reversed coded items. In this way, the remaining items loaded adequately to their respective constructs with good fit indices, i.e., CFI = 0.925, TLI = 0.915, RMSEA = 0.033, and SRMR = 0.060. In terms of internal consistencies, only two factors, namely leadership and communication displayed adequate in comparison with a cutoff value of 0.70. The factor loadings of the items and Cronbach's alpha for each construct are presented in Table 2.

Confirmatory Factor analysis (CFA) for the T-TPQ. The model fit for the T-TPQ displayed better fit indices than the T-TAQ with $\chi2(550) = 815.268$, $p < .001$ (normed chi-square = 1.48, falling within the recommended range from one to three), robust CFI = 0.884, TLI = 0.874, RMSEA = 0.053, and

Table 2. Factor loadings and Cronbach's alpha for each construct of the T-TAQ.

	Structure (α = 0.53)	Leadership (α = 0.75)	Monitor (α = 0.58)	Support (α = 0.55)	Communication (α = 0.72)
i1	0.517				
i2	0.441				
i3	0.354				
i4	0.321				
i5	0.383				
i6	0.349				
i7		0.667			
i8		0.515			
i9		0.635			
i10		0.522			
i11		0.602			
i12		0.555			
i13			0.463		
i14			0.521		
i15			0.424		
i16			0.398		
i17			0.370		
i18			0.398		
i19				0.659	
i22				0.436	
i23				0.529	
i24				0.397	
i25					0.568
i26					0.595
i27					0.608
i28					0.708
i29					0.562

SRMR = 0.063. The factor loadings identified three items with lower values than expected, including i1 (0.342), i2 (0.137), and i16 (0.273). Thus, we attempted to rerun the CFA by removing one item at a time, one of two the pairs at a time, and all three variables together to examine if all the remaining items could result in better fit and factor loadings. This entailed six post-hoc examinations of the model. The results of modified models after removing these items suggested a better CFA model. After examining the items semantically, we decided that the items may have little contribution to the respective constructs due to their semantic overlapping with others. Thus, the final CFA model for the T-TPQ consisted of 32 items with $\chi 2$ (451) = 643.289, $p < .0001$, robust CFI = 0.912, TLI = 0.903, RMSEA = 0.050, and SRMR = 0.061. Regarding internal consistencies, all factors displayed adequate reliability with Cronbach's alpha quite above the cutoff of 0.7. The factor loadings of the items and Cronbach's alpha for each construct are presented in Table 3.

Discriminant validity and post-hoc examination of factor structure

The average variance extracted (AVE) was calculated for the established factors from the CFA, which was then compared with the squared correlations among the constructs themselves. Table 6.S in the Supplementary material presents the AVEs diagonally and the squared correlations among the constructs. Accordingly, the results revealed discriminant validity issues with the two instruments such that most constructs did not obtain an AVE > 0.50 and their AVEs were lower than the squared correlations. Furthermore, the high correlations among the constructs suggested that they could not be strongly distinguished from each other.

On this basis, we also attempted the exploratory factor analysis (EFA) to explore the factor structure of the two instruments. As for the T-TAQ, the scree plot suggested either a one-factor or two-factor solution. The EFA showed that seven out of 30 items displayed low factor loadings on the first factor, which explained 18.9% variance. Only two negative items made up of the second factor. Altogether, the two factors explained 24.7% variance.

Regarding the T-TPQ, the scree plot revealed a two-factor solution, and the EFA confirmed this. Only leadership remained as a separate factor while all other items, except that items 2 and 16, loaded onto to first factor. Altogether, the two factors accounted for a variance of 34.9%. The factor loadings from these EFA can be found in Table 7.S in the Supplementary material.

Table 3. Factor loadings and Cronbach's alpha for each construct of the T-TPQ.

Loadings	Structure (α = 0.78)	Leadership (α = 0.90)	Monitor (α = 0.79)	Support (α = 0.76)	Communication (α = 0.76)
i3	0.587				
i4	0.698				
i5	0.554				
i6	0.718				
i7	0.649				
i8		0.763			
i9		0.736			
i10		0.731			
i11		0.754			
i12		0.771			
i13		0.837			
i14		0.717			
i15			0.624		
i17			0.725		
i18			0.715		
i19			0.547		
i20			0.649		
i21			0.56		
i22				0.461	
i23				0.449	
i24				0.596	
i25				0.696	
i26				0.504	
i27				0.587	
i28				0.453	
i29					0.639
i30					0.7
i31					0.616
i32					0.424
i33					0.514
i34					0.441
i35					0.589

Test-retest reliability

As can be seen in Table 4, the scores collected at two points in time displayed fair to good reliability with ICCs ranging from 0.720 and 0.798 for four constructs of the T-TAQ and ICCs from 0.704 to 0.843 for all constructs of the T-TPQ (Koo & Li, 2016). The leadership scale of the T-TAQ, with an ICC = 0.457, showed that there was certain fluctuation regarding the attitude toward leadership measured at a later time.

The SEM for the five factors constituting the T-TAQ ranged from 0.205 to 0.289. The minimum difference values at the individual levels within 0.569 and 0.802. This means in order to distinguish between a real change and measurement error, one participant should demonstrate an increase in the score that is greater than 0.569 for the situation monitoring, and 0.802 for the leadership scale of the T-TAQ.

For the T-TPQ, the SEM displayed higher values, ranging from 0.254 and 0.341, which were the expected deviations from the observed scores. Higher minimum difference values were found in four constructs of the T-TPQ in comparison with those from the T-TAQ. The highest minimum difference was 0.954 in the case of leadership and the lowest minimum difference was 0.705 as to the team structure.

Floor and ceiling effects

To examine the discriminative power of the T-TAQ and T-TPQ across the five constructs of team structure, leadership, situation monitoring, mutual support, and communication, frequencies of the lowest and highest factor scores were calculated. An average score was computed for each construct, ranging from 1 to 5 at maximum.

The result showed that a ceiling effect was found with more than 15% of the respondents obtaining the highest score of 5 for the leadership construct, despite no evidence of floor effect. Table 8.S in the Supplementary material displays the median and interquartile ranges (IQR) of each dimension as well as the percentage of participants reaching the maximum scores.

Group comparisons

Regarding the attitudes and perceptions of team work as measured by five constructs of the T-TAQ and T-TPQ, no

Table 4. Two-way ICCs absolute agreement for the T-TAQ and T-TPQ.

T-TAQ	ICC	Lower	Upper	SEM	Minimum difference
Team structure	0.720	0.612	0.801	0.279	0.772
Leadership	0.457	0.288	0.598	0.289	0.802
Situation monitoring	0.798	0.715	0.859	0.205	0.569
Mutual support	0.748	0.649	0.823	0.239	0.663
Communication	0.731	0.626	0.81	0.256	0.710
T-TPQ	ICC	Lower	Upper	SEM	Minimum difference
Team structure	0.735	0.631	0.813	0.254	0.705
Leadership	0.843	0.777	0.891	0.341	0.945
Situation monitoring	0.704	0.591	0.789	0.306	0.847
Mutual support	0.742	0.64	0.818	0.280	0.776
Communication	0.738	0.635	0.815	0.298	0.826

significant difference was observed between male and female participants.

Participants of different educational levels displayed differences in their perceptions toward teamwork only in the dimension of team structure. Pre-graduate degree holders had a higher median than bachelor's degree holders (p = .004) and post-graduate degree holders (p = .003). Table 9.S in the Supplementary material presented the results for group comparison for the five dimensions of the two questionnaires.

Discussion

In light of the wide-spread use of the T-TAQ and T-TPQ in French-speaking communities, we translated them and compared the back-translated versions with the original English versions based on three criteria, namely semantic equivalence, relevance, and ambiguity. As concerns the semantic equivalence, we found that these French versions could be considered as being both culturally relevant and semantically equivalent to the original English versions. Minor modifications resulted from the review process and the pilot testing, showing that certain concepts and word usage should be adapted according to the language itself and the target users, i.e., different synonyms were considered and finally chosen on the basis of the context of public health. Given the rigorous back-translation, panel reviewing, and two-round pilot testing of clarity, these French versions of the T-TAQ and T-TPQ can be considered as being appropriate for use by healthcare professionals working in health care units. To date, the T-TAQ and T-TPQ have been translated into different languages, e.g., Norwegian (Ballangrud et al., 2020), Swedish (Hall-Lord et al., 2021), Chinese (Qu et al., 2020), and Japanese (Unoki et al., 2021), which has demonstrated their great potential in teamwork training and assessment. The availability of the French version is deemed important to advance the usability of the tools and TeamSTEPPS® research at large.

The original English version (Baker et al., 2010) only examined the internal consistency for each of the five constructs of the T-TAQ, but not a CFA. We found Cronbach's alpha lower than those reported by Baker et al. (2010) with only leadership and communication being slightly higher the cutoff value of 0.70. Nevertheless, in comparison with a recent validation study of the Norwegian version of the T-TAQ (Ballangrud et al., 2020) and re-validation of the T-TAQ in school-based health teams (Wolk et al., 2020), certain similarities and differences were observed. First, while a five-factor structure of the T-TAQ was reserved in the Norwegian version, a unitary factor structure was more evident in the present study and in Wolk et al. (2020). Further analysis regarding the discriminant validity of the constructs indicated that between one pair of constructs of the T-TAQ, the correlations were so high that it was not possible to establish discriminant validity.

Moreover, the three reversed coded items (i20, i21, and i30) were found to be problematic such that sufficient loadings were not achieved. This is in line with the finding from Wolk et al. (2020) in which all four reversed coded items did not load on any factor. Research on reversed coded items seems to support that the inclusion of these items may not be recommended, because the interpretation between negatively and positively formulated items is not always equivalent, resulting in poor model fit (Marsh, 1986).

As concerns the T-TPQ, the present result from the CFA was mostly in line with the validation of the English version (Keebler et al., 2014); the five factors being confirmed with good model fit indices. The factor loadings suggested the removing of three items, two of which belonging to the team structure factor and one to the situation monitoring. Upon the examination of the items' meaning, an overlapping with other items in the same scale served as a possible explanation. For example, the meaning of item 2, read as "Staff monitor each other's performance", may have already been covered to item 21 "Staff correct each other's mistakes to ensure that procedures are followed properly". Thus, a parsimonious model with less items was preferred.

Due to the unfavorable results from the discriminant validity testing, our exploratory EFA suggested that only two factors for the T-TPQ were evident. This meant that a separation of the dimensions might not be clear-cut. The leadership scale could remain as a separate factor but not the others. With the same approach, Wolk et al. (2020) found a one-factor solution for the T-TPQ. The difference in the factor solution might result from our sample nature. In Keebler et al. (2014), the sample was large and consisted of a variety of functionalities such as administration, management, dental assistants, pharmacists, and registered nurses and in Wolk et al. (2020) mental health team members from public schools. Thus, for a particular target sample, the distinction between the factors may not be as important as the total score (Wolk et al., 2020).

As for internal consistency within one factor, the Cronbach's alpha for all the constructs, ranged from 0.76 to 0.90. Even if those figures are lower than the English version validation study by Keebler et al. (2014) or the Norwegian version (Ballangrud et al., 2017), except for the leadership scale, they confirm that internal consistency was established.

The scores collected at an interval of two to four weeks showed fair to good test-retest reliability for most of the T-TAQ items, and T-TPQ items. This reliability was lower than those found in Ballangrud et al. (2017) and Ballangrud et al. (2020). However, in the present study, the number of participants who repeated the measure was higher, 102 compared to 26 persons in Ballangrud et al. (2017) and Ballangrud et al. (2020). More importantly, the ICC reported here for repeated measures was two-way mixed effect, absolute agreement whereas in Ballangrud et al. (2017) and Ballangrud et al. (2020), two-way random ICCs were used. According to Koo and Li (2016), two-way mixed ICC is more appropriate for test-retest design. Thus, the differences may be attributable to methodological differences.

In general, with all ICCs above 0.7, except that from the leadership scale of the T-TAQ, test-retest reliability of the two instruments can be established (Terwee et al., 2007). The highest SEM and minimum difference value obtained for the attitude toward leadership mean that there were more variations between two measurements. This can be due to the measurement error or the actual change observed from the participants. In fact, there were 13 participants reporting a score exceeding the 95% confidence interval of the SEM. Additionally, one participant exposed a real change, i.e., the score at time 2 was 0.83, which was higher than the minimum difference. Therefore, further examination of both the test-retest reliability of the leadership scale of the T-TAQ to figure out the source of variation is suggested.

In line with other studies, e.g., Wolk et al. (2020) and Ballangrud et al. (2020), the scores of the attitudes and perceptions of teamwork across the five dimensions of the T-TAQ and T-TPQ were more positive, particularly the scores of the attitude dimensions. In this respect, Baker et al. (2010) found that without team training, the participants still obtained a positive attitude toward teamwork. Thus, it is difficult to measure a change in attitude as an outcome of team training because attitudes seem to be highly correlated with general intelligence (Baker et al., 2010; Stevens & Campion, 1999). This means that the two questionnaires, measuring teamwork knowledge and skills, require knowledge of problem solving and significant cognitive abilities rather than personal traits or dispositions (Stevens & Campion, 1999). For the above-mentioned reason of the limited range of the scoring of the T-TAQ and T-TPQ, almost no group differences as a function of gender and educational levels, except for the perception of team structure. This echoes findings from Baker et al. (2010) who found only a significant difference of less than 0.10 point between groups of different positions, which has little implication for possible interventions. Therefore, in order to initiate changes in teamwork attitudes and perceptions, teamwork training or team-based interventions should be coupled with managerial change effort, such as new technologies or quality enhancement, to enable individual team members learn and innovate (Stevens & Campion, 1999). This, in turn, could lead to changes in individuals' frame of references in dealing with newly encountered situations, which is a priori of attitude changes.

Based on the factor analysis, regarding the T-TAQ, we suggest that the three negative items (i20, i21, and i30) should be further examined in a different context as they did not load onto their expected factor. A two-factor structure was favored for the T-TPQ, of which the leadership dimension was retained. Given adequate test-retest reliability and limited range in the scoring, the two instruments of the TeamSTEPPS® can thus be used to measure attitudes and perceptions of teamwork in French-language institutions to define training needs. Findings from this study and from Baker et al. (2010) and Wolk et al. (2020) nevertheless suggest that use of the T-TAQ and T-TPQ as outcome measures should be reconsidered or only recommended within teams who score low at baseline.

In recent years and primarily in English-speaking countries, TeamSTEPPS® has proved its worth in improving teamwork performance, such as leadership and communication (Castner et al., 2012; Weaver et al., 2010), clinical management and outcomes (Capella et al., 2010; Forse et al., 2011; Neily et al., 2010; Weld et al., 2016), safety culture (Staines et al., 2020; Thomas & Galla, 2013) and work environment (Howe, 2014; Körner et al., 2015). However, the implementation of the TeamSTEPPS® program and tools remains scarce in French-speaking countries (Staines et al., 2020). The translation and validation of the T-TAQ and T-TPQ questionnaires may therefore facilitate future implementation and evaluation. Indeed, the literature reveals interesting varied results that may be partly attributable to cultural differences (Körner et al., 2015; Lakatamitou et al., 2020). Therefore, data collected in French-speaking research settings could expand TeamSTEPPS® research, enable comparisons with existing results, and thus help better understand how culture impacts the management and teamwork of healthcare teams.

Limitations

Several limitations should be acknowledged. The cross-sectional design and convenience sampling employed entail that generalization cannot be upheld. A sample size of 235 was adequate for confirmatory factor analysis such that a minimum of 5 cases were required for one item or a minimum sample size of 200 when there were many indicators for a latent variable (Bentler & Chou, 1987; Floyd & Widaman, 1995). Nevertheless, the participants came from only four institutions, working multiple care units. Therefore, future studies could try to involve a more heterogenous sample across different types of institution or outside the hospital context like in long-term care facilities. Additionally, taking into account the professions of the respondents would be valuable to understand the applicability of the two instruments in different team settings or team compositions. Moreover, to measure attitudes and perceptions, the use of self-reported measures means that social desirability bias could not be controlled for. Nevertheless, this aspect of social desirability was addressed during the development of the instrument by Baker

et al. (2010). Furthermore, when it comes to capturing attitudes and perceptions, self-reported measures are more efficient while objective measures are recommended for behavioral measurement.

Conclusions

By employing forward-backward translation approach and a rigorous review and adaptation process, we found that the resulting French-language versions of the T-TAQ and T-TPQ fulfill criteria of semantic equivalence and culturally relevance. High correlations among the different dimensions of the two instruments were found and test-retest reliability was established. In the context of public healthcare, we found that the use of T-TPQ total score when measuring attitudes and perceptions of teamwork is more appropriate. Additionally, more positive scores were found for the five dimensions of the T-TAQ.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors have no conflict of interest or financial disclosures to report.

Notes on contributors

Anh Nguyet Diep, obtained her PhD degree in Educational Sciences from the Vrije Universiteit Brussels in 2018 and master's degree of Statistics from the KU Leuven in the same year. Her research focuses on blended learning with students' learning and satisfaction, online interaction and instrument validation as major topics. Currently, she works as a biostatistician at the Biostats Unit, Department of Public Health, University of Liège.

Méryl Paquay, MSc, RNCC, RN: Mrs Paquay works as a quality and safety care manager. She is a nurse with a specialization in community health and a master's degree in public health. She is also certified to work as a simulation instructor. Mrs Paquay have recently started a PhD. Her area of expertise covers topics related to Crisis Resources Management, reporting, simulation and the impactof the work environment on the quality of care.

Jean-Christophe Servotte, RN, CCRN, MSc, PhD. Jean-Christophe Servotte is a critical care nurse. He holds a master's degree in public health science and a PhD in studying the impact of different simulation modalities on health learning. He is now working as researcher at Cellule Recherche & Développement Centre Namur Simulation, Haute Ecole Namur-Liège-Luxembourg. He is Chair of theINACSL International Committee.

Nadia Dardenne, MSc. is biostatistician at the Biostats Unit, Department of Public Health, University of Liège. She obtained a bachelor's degree in Mathematical Sciences and master's degree in Statistics from the University of Liège. Besides teaching, her research activities consist of providing methodological and statistical support to numerous national and international projects.

Laure Istas holds a master's degree in political science, International Relations with a focus on humanitarian action. She worked at the FPS Public Health, Safety of the Food Chain and Environment within the Quality and Patient Safety Unit. She has acquired knowledge and expertise in the field of coordination of quality of care and patient safety and has worked with hospitals at the national level to publish recommendations on different themes. Laure is now a project manager for the PAQS and works on various projects, focusing on patient involvement and participation, teamwork and the MR/MRS component.

Pauline Van Ngoc obtained her bachelor's and master's degrees in Industrial and Organizational Psychology at the University of Liège. She now works at the Unité de Recherche Soins Primaires et SantéDépartement de Médecine Générale, University of Liège,Liège, Belgium.

Mathilde Charlier obtained her master's degree in Public Health, University of Liège in 2019. She is currently working at the Department of Public Health, University of Liège. Teamwork skills and training are her research topics of interest.

Alexandre Ghuysen, MD, PhD (male) qualified as medical doctor from the Liège University in 1992, trained in internal medicine in 1997 and emergency medicine in 1999, and PhD in Biomedical and experimental science in 2006. He has been the promoter of several PhD thesis, notably in the field of emergency and acute care medicine.At present, he is both professor at Liège University where he teaches acute care medicine, Head of the Center for Medical Simulation at the same University and associated Head of the Emergency Department at CHU Liège. He is currently the President of the Belgian Society of Emergency and Disaster Medicine (BESEDIM). He has received funding for several national and international research projects (Interreg).

Anne-Françoise Donneau studied mathematics at the University of Liège and graduated in 2002. In 2013, she passed her PhD thesis on incomplete longitudinal ordinal data at the University of Liège. In 2015, she was appointed full professor of biostatistics at the Faculty of Medicine at the University of Liège and Director of the Biostatistics and Research methods Unit. Since 2016, she serves as medical computing expert for the steering committee within the framework of breast screening, Ministry of Public Health. She is currently involved as statistician expert in several INTERREG and regional projects. Her research interests include ordinal variables, missing data, multiple imputation, longitudinal/correlated data, cluster methods and validation techniques.

ORCID

Anh Nguyet Diep http://orcid.org/0000-0001-5134-3220
Anne-Françoise Donneau http://orcid.org/0000-0002-5385-6289

References

Baker, D., Salas, E., Barach, P., Battles, J., & King, H. (2007). The relation between teamwork and patient safety. In Carayon P. (ed.): *Handbook of human factors and ergonomics in health care and patient safety*. Mahwah, NJ: Lawrence Erlbaum (pp. 259–271).

Baker, D. P., Amodeo, A. M., Krokos, K. J., Slonim, A., & Herrera, H. (2010). Assessing teamwork attitudes in healthcare: Development of the TeamSTEPPS teamwork attitudes questionnaire. *Quality & Safety in Health Care*, *19*(6), 1–4. DOI:10.1136/qshc.2009.036129.

Baker, D. P., Krokos, K. J., & Amodeo, A. M. (2008). *TeamSTEPPS teamwork attitudes questionnaire manual*. Washington, DC: American Institutes for Research.

Baker, D. P., Salas, E., King, H., Battles, J., & Barach, P. (2005). The role of teamwork in the professional education of physicians: Current status and assessment recommendations. *The Joint Commission Journal on Quality and Patient Safety*, *31*(4), 185–202. https://doi.org/10.1016/S1553-7250(05)31025-7

Ballangrud, R., Husebø, S. E., & Hall-Lord, M. L. (2017). Cross-cultural validation and psychometric testing of the Norwegian version of the TeamSTEPPS® teamwork perceptions questionnaire. *BMC Health Services Research*, *17*(1), 799. https://doi.org/10.1186/s12913-017-2733-y

Ballangrud, R., Husebø, S. E., & Hall-Lord, M. L. (2020). Cross-cultural validation and psychometric testing of the Norwegian version of TeamSTEPPS teamwork attitude questionnaire. *Journal of*

Interprofessional Care, *34*(1), 116–123. https://doi.org/10.1080/13561820.2019.1638759

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Brandão, C. S., Fernandes, D. C., Collares, C. F., Fernandes, G. R., & Marin, H. F. (2016). Tradução e retrotradução do instrumento de avaliação do trabalho em equipe TeamSTEPPS® para uso no ensino por simulação no Brasil. *Scientia Medica*, *26*(4), 13. https://doi.org/10.15448/1980-6108.2016.4.24622

Capella, J., Smith, S., Philp, A., Putnam, T., Gilbert, C., Fry, W., Harvey, E., Wright, A., Henderson, K., Baker, D., Ranson, S., & ReMine, S. (2010). Teamwork training improves the clinical care of trauma patients. *Journal of Surgical Education*, *67*(6), 439–443. https://doi.org/10.1016/j.jsurg.2010.06.006

Castner, J., Foltz-Ramos, K., Schwartz, D. G., & Ceravolo, D. J. (2012). A leadership challenge: Staff nurse perceptions after an organizational TeamSTEPPS initiative. *Journal of Nursing Administration*, *42*(10), 467–472. https://doi.org/10.1097/NNA.0b013e31826a1fc1

Core Team, R. (2013). R: A language and environment for statistical computer. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org

Flin, R. H., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end: A guide to non-technical skills*. Ashgate Publishing, Ltd.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286–299. https://doi.org/10.1037/1040-3590.7.3.286

Forse, R. A., Bramble, J. D., & McQuillan, R. (2011). Team training can improve operating room performance. *Surgery*, *150*(4), 771–778. https://doi.org/10.1016/j.surg.2011.07.076

Hall-Lord, M. L., Bååth, C., Ballangrud, R., & Nordin, A. (2021). The Swedish version of the TeamSTEPPS® teamwork attitudes questionnaire (T-TAQ): A validation study. *BMC Health Services Research*, *21*(1), 1–8. https://doi.org/10.1186/s12913-021-06111-1

Harlow, L. L. (2014). *The essence of multivariate thinking: Basic themes and methods*. Routledge.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*(1), 1–15. https://doi.org/10.2165/00007256-200030010-00001

Howe, E. E. (2014). Empowering certified nurse's aides to improve quality of work life through a team communication program. *Geriatric Nursing*, *35*(2), 132–136. https://doi.org/10.1016/j.gerinurse.2013.11.004

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hwang, J.-I., & Ahn, J. (2015). Teamwork and clinical error reporting among nurses in Korean hospitals. *Asian Nursing Research*, *9*(1), 14–20. https://doi.org/10.1016/j.anr.2014.09.002

Keebler, J. R., Dietz, A. S., Lazzara, E. H., Benishek, L. E., Almeida, S. A., Toor, P. A., King, H. B., & Salas, E. (2014). Validation of a teamwork perceptions measure to increase patient safety. *BMJ Quality & Safety*, *23*(9), 718–726. https://doi.org/10.1136/bmjqs-2013-001942

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). Institute of Medicine (US) Committee on Quality of Health Care in America. In *To err is human: Building a safer health system*. National Academies.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Körner, M., Wirtz, M. A., Bengel, J., & Göritz, A. S. (2015). Relationship of organizational culture, teamwork and job satisfaction in interprofessional teams. *BMC Health Services Research*, *15*(1), 243. https://doi.org/10.1186/s12913-015-0888-y

Lakatamitou, I., Lambrinou, E., Kyriakou, M., Paikousis, L., & Middleton, N. (2020). The Greek versions of the TeamSTEPPS teamwork perceptions questionnaire and Minnesota satisfaction questionnaire "short form". *BMC Health Services Research*, *20*(1), 587. https://doi.org/10.1186/s12913-020-05451-8

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

MacKenzie, S. B., & Podsakoff, P. M. (2012). Common method bias in marketing: Causes, mechanisms, and procedural remedies. *Journal of Retailing*, *88*(4), 542–555. https://doi.org/10.1016/j.jretai.2012.08.001

Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: A review of the literature. *Acta Anaesthesiologica Scandinavica*, *53*(2), 143–151. https://doi.org/10.1111/j.1399-6576.2008.01717.x

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, *22*(1), 37. https://doi.org/10.1037/0012-1649.22.1.37

Mayer, C. M., Cluff, L., Lin, W.-T., Willis, T. S., Stafford, R. E., Williams, C., Saunders, R., Short, K. A., Lenfestey, N., Kane, H. L., & Amoozegar, J. B. (2011). Evaluating efforts to optimize TeamSTEPPS implementation in surgical and pediatric intensive care units. *The Joint Commission Journal on Quality and Patient Safety*, *37*(8), 365–374. https://doi.org/10.1016/S1553-7250(11)37047-X

Neily, J., Mills, P. D., Young-Xu, Y., Carney, B. T., West, P., Berger, D. H., Mazzia, L. M., Paull, D. E., & Bagian, J. P. (2010). Association between implementation of a medical team training program and surgical mortality. *Jama*, *304*(15), 1693–1700. https://doi.org/10.1001/jama.2010.1506

Peberdy, M. A., Kaye, W., Ornato, J. P., Larkin, G. L., Nadkarni, V., Mancini, M. E., Nichol, G., Lane-Trultt, T., & R A., B. (2003). Cardiopulmonary resuscitation of adults in the hospital: A report of 14 720 cardiac arrests from the National Registry of Cardiopulmonary Resuscitation. *Resuscitation*, *58*(3), 297–308. https://doi.org/10.1016/S0300-9572(03)00215-6

Qu, J., Zhu, Y., Cui, L., Yang, L., Lai, Y., Ye, X., & Qu, B. (2020). Psychometric properties of the Chinese version of the TeamSTEPPS teamwork perceptions questionnaire to measure teamwork perceptions of Chinese residents: A cross-sectional study. *BMJ Open*, *10*(11), e039566. https://doi.org/10.1136/bmjopen-2020-039566

Salas, E., Sims, D. E., & Burke, C. (2005). Is there a "Big Five" in teamwork? *Small Group Research*, *36*(5), 555–599. https://doi.org/10.1177/1046496405277134

Silva, R., Saraiva, M., Cardoso, T., & Aragão, I. C. (2016). Medical Emergency Team: How do we play when we stay? Characterization of MET actions at the scene. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *24*(1), 1–6. https://doi.org/10.1186/s13049-016-0222-7

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. https://doi.org/10.1111/j.1365-2753.2010.01434.x

Staines, A., Lécureux, E., Rubin, P., Baralon, C., & Farin, A. (2020). Impact of TeamSTEPPS on patient safety culture in a Swiss maternity ward. *International Journal for Quality in Health Care*, *32*(9), 618–624. https://doi.org/10.1093/intqhc/mzz062

Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, *25*(2), 207–228. https://doi.org/10.1177/014920639902500205

Terwee, C., Bot, S., De Boer, M., Van Der Windt, D., Knol, D., Dekker, J., Bouter, L., & De Vet, H. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Thomas, L., & Galla, C. (2013). Building a culture of safety through team training and engagement. *BMJ Quality & Safety*, *22*(5), 425–434. https://doi.org/10.1136/bmjqs-2012-001011

Unoki, T., Matsuishi, Y., Tsujimoto, T., Yotsumoto, R., Yamada, T., Komatsu, Y., Kashiwakura, D., & Yamamoto, N. (2021). Translation, reliability and validity of Japanese version the TeamSTEPPS® teamwork perceptions questionnaire. *Nursing Open*, *8*(1), 115–122. https://doi.org/10.1002/nop2.609

Weaver, S. J., Rosen, M. A., DiazGranados, D., Lazzara, E. H., Lyons, R., Salas, E., Knych, S. A., McKeever, M., Adler, L., Barker, M., & King, H. B. (2010). Does teamwork improve performance in the operating room? A multilevel evaluation. *The Joint Commission Journal on Quality and Patient Safety*, *36*(3), 133–142. https://doi.org/10.1016/S1553-7250(10)36022-3

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, *19*(1), 231–240. https://pubmed.ncbi.nlm.nih.gov/15705040/

Weld, L. R., Stringer, M. T., Ebertowski, J. S., Baumgartner, T. S., Kasprenski, M. C., Kelley, J. C., Cho, D. S., Tieva, E. A., & Novak, T. E. (2016). TeamSTEPPS improves operating room efficiency and patient safety. *American Journal of Medical Quality*, *31*(5), 408–414. https://doi.org/10.1177/1062860615583671

Winters, B. D., Weaver, S. J., Pfoh, E. R., Yang, T., Pham, J. C., & Dy, S. M. (2013). Rapid-response systems as a patient safety strategy: A systematic review. *Annals of Internal Medicine*, *158*(5_Part_2), 417–425. https://doi.org/10.7326/0003-4819-158-5-201303051-00009

Wolk, C. B., Locke, J., Salas, E., Eiraldi, R., Cronholm, P. F., & Mandell, D. (2020). An examination of the factor structure of TeamSTEPPS measures in school mental health teams. *Journal of Psychologists and Counsellors in Schools*, *30*(2), 172–184. https://doi.org/10.1017/jgc.2019.18