

Article

---

# The Performance of the Health Communication Assessment Tool© (HCAT-f) in Calibrating Different Levels of Nurse Communication Skills in a French-Speaking Context

## Evidence and Clinical Implications

DOI: 10.47368/ejhc.2022.308  
2022, Vol. 3(3) 164-179  
CC BY 4.0

**Anh Nguyet Diep** 

Public Health Sciences Department, University of Liège, Belgium;  
Department of Information Technology, Can Tho University, Vietnam

**Jean-Christophe Servotte** 

Paramedical Department, Namur-Liège-Luxembourg University College, Belgium

**Nadia Dardenne**

Public Health Sciences Department, University of Liège, Belgium

**Sophie Vanbelle** 

School for Public Health and Primary Care, Maastricht University, the Netherlands

**Vanessa Wauthier**

Centre Hospitalier Régional de Namur, Belgium

**Méryl Paquay**

Medical Simulation Center of Liege, University of Liège, Belgium

**Suzanne Hetzel Campbell** 

Department of Applied Science, School of Nursing, The University of British Columbia,  
Canada

**Jonathan Goffoy**

Medical Simulation Center of Liege, University of Liège, Belgium

---

**Corresponding author:**

Anh Nguyet Diep: [anhnguyet.diep@uliege.be](mailto:anhnguyet.diep@uliege.be)

**Anne-Françoise Donneau** 

Public Health Sciences Department, University of Liège, Belgium

**Alexandre Ghuyesen**Public Health Sciences Department, University of Liège, Belgium;  
Medical Simulation Center of Liege, University of Liège, Belgium**Abstract**

Communication skills training is essential in nurse education. Miscommunication may lead to adverse events and unsafe healthcare. To date, valid and reliable instruments to serve both communication training and assessment purposes across different cultural contexts are scarce. The present study empirically tested a French-language version of the Health Communication Assessment Tool© (HCAT-f) across different levels of communication skills performance to establish its reliability and validity through a cognitive fluency framework. Ten experts in communication and 52 nurse educators rated three videos simulating conversations between a nurse and a patient scheduled for lumpectomy. Each video captured a different level of communication skills performed by the nurse: High, medium, and low. Three distinct constructs were identified, i.e., professional presentation, empathy, and trust building. At absolute single-measure, an ICC = .43 suggested adequate interrater reliability of the whole scale for the medium-performed scenario, which decreased in low-performed (ICC = .35) and high-performed (ICC = .18) scenarios. The HCAT-f fulfils the criteria of linguistic equivalence, contextual relevance, and demonstrates acceptable construct validity. It can be used as a summative assessment tool after prior training on scale calibration is in place because interrater agreement was difficult to be established in high and low performance scenarios.

**Keywords**

Healthcare, communication skills, cognitive fluency, scale validation.

Nurse-patient communication has played a central role in determining the quality of nursing care, patients' well-being, and satisfaction (Williams & Gosette, 2011). Indeed, Fleischer et al. (2009) proposed that the purpose of nurse-patient communication should be to enhance patients' health and well-being. Ineffective communication might result in preventable medication errors, delayed treatment, misdiagnosis, or patients' injuries (Foronda et al., 2016), making it mandatory to increase nurses' communication skills and competences. Rather than an attempt to change communication, Kerr et al. (2022) argued that nurse-patient communication must be improved. Education and skills training/practice are central to behaviour change. Reaching that goal involves the ability to evaluate nurse-patient communications properly through a rigorous and valid instrument (Hargie et al., 1998). Assessment, as an integral part of education programs, focuses on the course objectives and enhances our understanding of how learning takes place and whether knowledge has transferred to skill development (Biggs, 2003).

Communication between nurses and patients is a dynamic process, affected by an array of factors, e.g., the nurse' motivation and cultural background, the patient's health status and life circumstances, and time constraints (O'Halloran et al., 2012). In view of the importance of the

regional and cultural aspects in communication, the reliability and validity of the assessment tools should also be seen through these different socio-cultural perspectives (Kardong-Edgren et al., 2010). Several aspects, such as the verbal and non-verbal forms used by the healthcare practitioner to express the message content and its emotional dimension should also be considered. From a cognitive fluency perspective, the ease by which certain traits or characteristics can be perceived might affect their judgement (Alter & Oppenheimer, 2009). Accordingly, some people will rate a trait as being present or truthful more easily if it is more noticeable.

In a review of current existing instruments used to teach and evaluate nurses' communication skills, Campbell et al. (2017) noticed that there was a lack of valid and reliable instruments which can be used in the context of simulation-based teaching. Most instruments either focus on a specific type of therapeutic relationship, e.g., with patients with cancer (Aranda & Yates, 2009), or communication skills like jargon explanation (Caron et al., 2013) or from a patient perspective (T-com skill scale, Baumaan et al., 2008).

The Health Communication Assessment Tool (HCAT) was developed to provide an instrument for the faculty and the students to assess the communication skills in a simulation-based education context (Campbell et al., 2013). This tool built around five distinct constructs, i.e., empathy, introduction, trust building, patient-family education, and power sharing has demonstrated good discriminant validity (Pagano et al., 2015). However, the validation process was conducted on a heterogeneous international sample in English, Portuguese and Korean languages, limiting per se its generalisation to populations that differ in language, social or cultural background (Baird et al., 2021; Goes et al., 2017; Pagano et al., 2015; Reis et al., 2018; Yang et al. 2016). Furthermore, test-retest reliability was not addressed, which limits insights to the practical applicability.

In the present study, we aimed to validate a French-language version of HCAT (HCAT-f) on a sample of nurse educators in Belgium, who were the ultimate users of the instrument. What distinguishes the study from previous research is the development of three simulated videos capturing different levels of nurses' communication competences. Further, two important psychometric properties of the HCAT, namely the interrater and test-retest reliability, were thoroughly evaluated.

## Method

### *French-Language Version of the HCAT: The Translation Procedure*

First, the 22 items of the HCAT (Pagano et al., 2015) were translated from English to French with the forward-backward translation method (Bullinger et al., 1993) by two independent French speakers who were competent in both the English and French languages as well as having expertise in the field of nursing education and communication. Next, a committee including the two faculty members, a researcher in public health with a nursing background, specialised in simulation and a communication expert with a PhD in psychology also specialised in simulation, and the same two professional translators, from the forward-backward translation, reviewed the two French translations. This helped to provide a more objective and accurate assessment of the draft versions (Vallerand, 1989). Where discrepancies were found, examination of the English original version made it possible to choose the most appropriate French translation. To this end, the committee members used a systematic strategy

by assessing each word to have the closest meaning to the English version. The final version was validated through consensus among the committee members.

### *Content Validity*

The content validity of the HCAT-f was captured by expert ratings of relevance at two time points. First, 10 experts (nine nurse educators and one psychologist and nurse educator were invited to rate the relevance of the HCAT-f items on a seven-point scale from 1 (*not relevant*) 1 to 7 (*very relevant*). The nurse educators taught nurse-patient communication in two nursing schools in Namur (Belgium). The psychologist worked in the University of Liège and specialised in therapeutic communication. All the experts were females with a mean age of  $40.90 \pm 8.06$  and  $14.50 \pm 10.48$  years of experience working in their respective field. Out of these 10 experts, five who showed both availability and enthusiasm were invited to rate the relevance of the items in a second round. The item-level content validity index (I-CVI), scale level S-CVI/UA (universal agreement method), and S-CVI/AVE (average method) were calculated to decide if any item had to be removed.

In the first rating round, the S-CVI/UA was .09 and the S-CVI/AVE .79 with items 2 and 13 having the lowest I-CVI of .50. Additionally, items 8, 9, 10, 15, and 21 did not meet the threshold of .78 (Polit et al., 2007). Based on consultations, item 2 was removed due to cultural irrelevance and item 13 was modified. A precision “if needed” was added for item 10. In the second round, there was an improvement after the modification. The S-CVI/UA was .86 and the S-CVI/AVE .96, which revealed adequate scale content validity. All the items also demonstrated sufficient content validity with I-CVI ranging from .80 to 1.00. The result can be found in Tables A.1 and A.2 in the Appendix.

### *Simulation Videos*

Three videos were developed by one of the co-authors, capturing the interaction between a 25-year-old female patient and a 33-year-old female nurse. The three videos showed the same clinical case but with communication performances that were either poor, medium, or highly competent. The clinical case involved a nurse taking the vital signs of a patient in a surgical unit who needed a general anaesthetic and a lumpectomy. The scripts for these scenarios were developed based on the HCAT-f items. Accordingly, video 1 (3:20 minutes) demonstrated a situation where the nurse was highly competent in her communication performance. Video 2 (2:05 minutes) showed a situation in which the performance was above average but could not be considered as excellent; and video 3 (1:30 minutes) showed communication skills at their minimum.

### *Procedure*

After the study had been reviewed and approved by the university ethics committee (reference number 2017/256 on October 17, 2017), the data collection process started between January and April 2018. The participants ( $N = 52$ ) were nurse educators from 15 nursing schools in the French-speaking part of Belgium. To reach a good number of responses, convenience sampling was employed.

They were asked to rate the three simulated videos against the HCAT-f items twice, with a four-week interval. To reduce any memory effect, the order in which videos were displayed

was randomised. The time given for watching the videos and filling the HCAT was set at 30 minutes to avoid fatigue and bias.

### *Referential Ratings*

Three experts in communication including one psychologist and two nurse educators who were involved in the referential rating evaluated the communication skills of the nurse in the simulated videos. A reference standard was defined as the mean of experts' ratings. The participants' ratings were then compared with this reference to establish the criterion validity (Terwee et al., 2007).

### *Data analysis*

**Construct Validity and Reliability.** To validate the factor structure or construct validity of the HCAT-f, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted. Oblique rotation was chosen instead of varimax because the latent constructs were expected to be correlated (Field, 2013). Prior to EFA, the wave 1 and 2 datasets were tested for fulfilment of factor analysis requirements. Kaiser–Meyer–Olkin (KMO) and Bartlett's test of sphericity (Pallant, 2010) were checked. Once these requirements were satisfactory, the number of factors extracted was based on the eigenvalues being greater than one (Bandalos & Boehm-Kaufman, 2009), the screeplot and the factor loadings ( $> .30$ ).

Next, the measurement model consisting of the factors and their respective observed variables were tested by CFA. We treated the scaling as being ordinal and weighted Least Squares (WLS) estimation was chosen (Byrne, 1998). The model was evaluated against the goodness-of-fit indices: Comparative fit index (CFI) and Tucker-Lewis index (TLI) above .90, Root Mean Square Error of Approximation RMSEA ( $< .06$ ), and standardised root mean square residual SRMR ( $< .08$ ; Hu & Bentler, 1999).

Finally, Cronbach's alpha was calculated to examine the internal consistency of the established factors.

### *Test-Retest Reliability*

To evaluate the stability of the rating scores obtained after a period of four weeks, namely test-retest reliability, two-way fixed intraclass correlation coefficient (ICC), absolute agreement was conducted for the three videos at both item and factor levels.

### *Interrater Reliability*

Two-way random ICC was run at the item and the factor levels. Both the consistency and absolute ICCs at average and single measures ICCs were reported to demonstrate any discrepancies if any. This will help to determine how many educators are needed to get reliable scores and therefore to establish the practical use of the HCAT-f.

### *Criterion Validity*

The ratings from one participant were compared with those from each of the three experts per category. Subsequently, the mean of these deviations was derived to reveal how participants and experts differed in their ratings across the three videos. To graphically display the distance between participants and experts' ratings, Bland-Altman plots were presented.

## Results

Raters ( $N = 52$ ) were nurse educators working at 15 nursing schools in Belgium. Most of the participants (82.7%) were females. Roughly two-thirds of the participants were Belgian, and another one-third was of French nationality with an average of 14.3 years of experience ( $SD \pm 9.4$  years). No missing data were observed. However, ratings of respondent 4 were removed because there was no variability in the item ratings. Normality was examined by means of skewness and kurtosis, which revealed acceptable cut-off values ( $< |2.00|$ ).

### *Construct Validity and Reliability*

**Exploratory Factor Analysis.** The KMO values were .96 and .97 and significant Bartlett's test of sphericity for wave 1 and wave 2 data, respectively. The sample size and communalities were therefore sufficient to proceed with factor analysis. Based on Pagano et al. (2015), the first EFA was run with a fixed number of five factors to be extracted using oblique oblmin. The result indicated that there were only two items loading onto factors 4 and 5. All factor loadings were quite above the cut-off of .30. Three factors with eigen values greater than one were evident and the scree plot suggested a three-factor solution.

A second EFA was run with a fixed number of three factors. The result supported the three-factor structure with all loadings greater than .32. Altogether, the three factors explained a total variance of 63.80%.

The established factors can be explained as follows. First, factor one consists of items specifying behaviours regarding nurses' self-introduction, how they control their use of tone and jargon, and transfer of medical knowledge. Thus, we label this factor as "professional presentation". The second factor comprises items relating to the handling of the conversation in a fair manner and reducing the power difference between the nurse and the patient; consequently, it is assigned as "empathy" to be in line with Pagano et al. (2015). The third factor contains items related to the effort to enhance the proximity between the nurse and the patient. Therefore, we attribute this factor as "trust building", to be consistent with Pagano et al. (2015). The result can be found in Table A.3 in the Appendix.

**Confirmatory Factor Analysis.** A model with three identified factors consisting of 20 items was subject to a CFA using wave 2 dataset. All item loadings were greater than .60. Good model fit was achieved with  $CFI = .99$ ,  $RMSEA = .58$  [95%CI: .04; .07], and  $SRMR = .41$ . The factor loadings and Cronbach's alpha for each latent construct are presented in Table 1.

### *Test-Retest Reliability*

To evaluate if the ratings from the same raters of a particular item remained stable over time, two-way fixed ICC, absolute agreement was employed. The derived results are presented in Table A.4 in the Appendix. In general, the findings displayed poor test-retest reliability for most items in video 1, with only items 1, item 18, and item 20 demonstrating fair reliability. The test-retest reliability improved in the case of video 2, i.e., more than half of the items displayed fair to good reliability. Concerning video 3, only nine out of 21 items demonstrated fair to good test-retest reliability. At factor level, two of three factors demonstrated fair test-retest reliability in video 2 whereas poor test-retest reliability were observed for videos 1 and 3.

**Table 1.** Factor Loadings from CFA and Reliability Coefficients (Cronbach's Alpha) of the Latent Constructs

Items	F1 Professional presentation ( $\alpha = .92$ )	F2 Empathy ( $\alpha = .97$ )	F3 Trust building ( $\alpha = .67$ )
i1 Introduced herself/himself to the patient (and/or family).	.83		
i3 Explained the reason for her/his visit in appropriate terms.	.84		
i5 Maintained eye contact when talking with the patient and/or family.	.92		
i6 Communicated what s/he was about to do prior to doing it.	.82		
i13 Effectively educated the patient and/or family member about the procedure, disease, and/or treatment.	.88		
i16 Used appropriate vocal tone and volume for the situation.	.86		
i22 Avoided healthcare jargon (vital signs, respiratory, etc.)	.70		
i4 Used positive communication including a smile to encourage interactions.		.93	
i7 Asked the patient or family member if it was okay to touch the patient before doing a procedure or test (blood pressure, auscultation, IV, nasogastric intubation, etc.)		.64	
i10 Sat when talking with or educating the patient and/or family member.		.91	
i11 Listened more than talked.		.92	
i12 Leaned toward the patient or family member, who was speaking.		.92	
i14 Asked questions to encourage feedback and enhance clarity.		.87	
i15 Recognized and responded appropriately to the patient's and/or family member's nonverbal (frowns, tears, hysteria, silence, etc.) and verbal behaviours.		.89	
i18 Spent equal or more time on psychosocial aspects of patient/family care as on clinical (biological) aspects.		.90	
i19 Inquired about the patient's/family member's feelings regarding the situation.		.93	
i20 Recognized conflict and tried to gain information and found opportunities to minimise or manage it		.87	
i21 Maintained, enhanced, or developed an interpersonal relationship with the patient and/or family member (via communication and professionalism)		.95	
i8 Touched the patient appropriately.			.63
i9 Spent most time near the patient.			.63
i17 Avoided judging patient/family behaviours (re: economic status, abuse, drug use, sexual orientation, religion/cultural differences, etc.).			.63

Note.  $\alpha$  = Cronbach's alpha.

### *Interrater Agreement*

At item level, excellent interrater reliability was found for average measures across three videos, i.e., absolute measures ICCs ranged from .92 to .98 with video 2 revealing the highest interrater agreement (ICC = .98). Adequate interrater reliability was only found in the ratings of scenario 2 with absolute single measure ICC = .43 [.30; .61] whereas the result showed poor interrater agreement for both for video 1 (ICC = .18 [.10; .32]) and video 3 (ICC = .35 [.23; .53]). Spearman Brown prophecy formula (Spearman, 1910) was employed to determine the number of raters needed to achieve a good average single measure of .60. The result revealed that we would need roughly seven raters for the ratings for video 1, two for video 2, and three for video 3.

At factor level, regarding average measures, the ratings for the three videos displayed excellent interrater agreement with ICCs being higher than .90. At single measures, the results showed that good interrater agreement were found for video 2 (ICC = .65 [.32; .99]) and video 3 (ICC = .63 [.31; .99]). For video 1, the absolute single measure was poor (ICC = .20 [.06; .91]). To obtain a good absolute single measure for video at factor level, six raters were required. The results can be found in Tables A.5 and A.6 in the Appendix.

### *Criterion Validity*

The ratings of the experts correspond to the skills demonstrated in the video such that they gave high ratings for video 1, fair ratings for video 2, and low ratings for video 3. Based on the mean absolute deviations calculated for one participant and one expert at a time, the highest value was found in video 2. The absolute distance between the ratings of participants and experts at group level is 0.79 point on a 5-point scale (95%CI: 0.69; 0.91) for video one, 1.31 (95%CI: 1.12; 1.54) for video 2, and 0.87 (95%CI: 0.77; 0.96) for video 3.

The differences between participants and the reference rating were depicted in Bland-Altman plots (Figure 1). In these plots, the limits of agreements (LoA) are expected to contain 95% of the differences between the ratings of participants and the reference rating. Accordingly, the average of mean difference was the lowest in video 3 (bias = -0.11 [95%CI: -0.26; 0.03], LoA = -0.74, 0.51), followed by video 1 (bias = -0.18 [95%CI: -0.31; 0.04], LoA = -0.76, 0.41). Regarding video 2, a high bias = -0.88 [95%CI: -1.07; -0.68] was observed with LoA = (-1.72, -0.03). In other words, there was a systematic bias in the ratings of participants and experts in the first two videos. Thus, if we set a maximum of  $\pm 1$  point as acceptable difference between participants and experts' ratings, this criterion was only achieved in the case of low-performance, i.e. video 3.



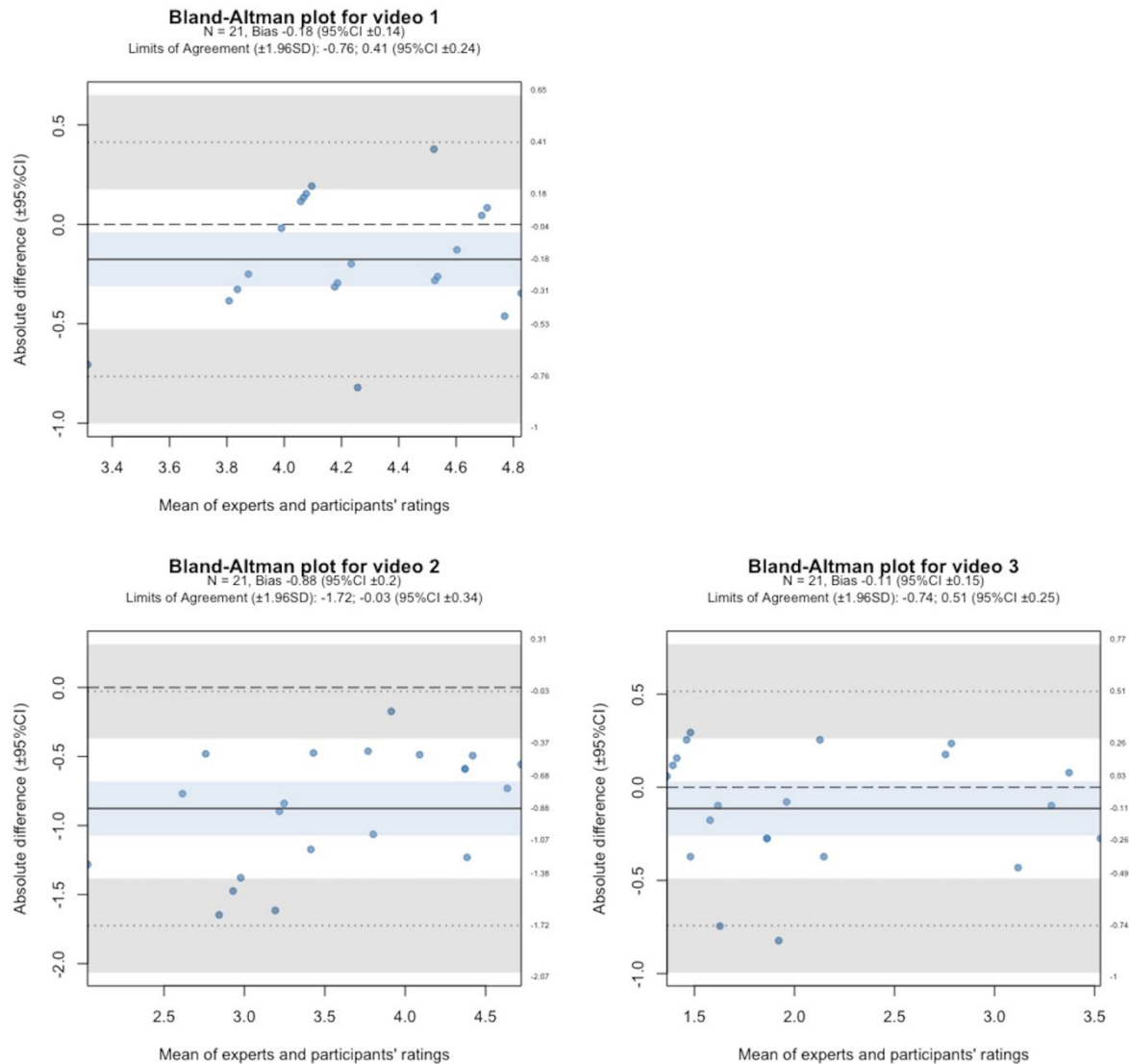


Figure 1. The Differences Between the Participants' and Experts' Ratings for the Three Videos

## Discussion

The present study aimed to validate the HCAT-f using videos simulating nurse-patient interaction at different levels of communication skills performance on a sample of nurse educators in Belgium. The findings are discussed in light of comparison with the English-version and cognitive fluency theory.

Regarding content equivalence, the translation of the HCAT-f suggested removing item 2, namely “shake the patient or family’s hand” due to cultural irrelevance. All other items making up the instrument could be semantically equivalent to their counterparts in the English version via the back-translation process. Additionally, expert ratings of relevance also demonstrate adequate relevance based on the I-CVI. Therefore, at face-validity, the HCAT can fulfil the criteria of linguistic equivalence and contextual relevance.

By means of EFA and CFA, a different factor structure from the original studies (Campbell et al., 2013; Pagano et al., 2015) emerged. A three-factor was more evident than the five-factor

solution. Particularly, the empathy and power sharing from the original scale now collapsed into one factor. Similar tendency was also observed for the other two factors of introduction and patient-family education. Only trust building remained as a separate factor. Based on the results and the three-factor structure recently endorsed by Baird et al. (2021), it is suggested that a more refined theoretical framework to conceptualise health communication competencies is of significance. In this regard, this validation study, together with recent works on the health communication skills assessment and training (e.g., Chen et al., 2021; Mendi et al., 2020), provides a strong empirical basis. From a statistical point of view, Jung and Lee (2011) suggest that the small sample size may result in the instability in the factor structure. This can be the case in both Campbell et al. (2013) and the current study, of which the sample sizes just fulfil the minimum requirement. In terms of internal consistency, two of the three constructs illustrated acceptable Cronbach's alpha, which is in line with results from Campbell et al. (2013).

The interrater agreement absolute single measures was low whereas excellent interrater reliability for average measures was achieved. This means the ratings by a single nurse educator is not reliable whereas the average rating of the whole sample is. This is in line with findings from Pagano et al. (2015). Based on cognitive fluency theory, we examined if there were differences in interrater reliability across three scenarios in which high, medium, and low performance of communication skills of the nurses were exhibited. We argued that when the traits were easily observed, interrater reliability would be higher. The results illustrate this perspective to some extent. More specifically, when the nurses demonstrated a lack of the skills or when their skills were not up to a high standard (scenarios 2 and 3), the agreement among the raters was higher. Reversely, the mean scores in scenario 1 is higher than the other two, but interrater agreement is the lowest. The complexity of the assessment process requires higher cognitive load such that the raters need to be highly focused to give a 4 or a 5 to a complex item like 'spent equal or more time on psychological aspects of the patient as well as on clinical (biological) aspects. Thus, when more nuances were important to inform our evaluation of a nurse's communication skill, more raters ( $n = 7$ ) are needed. Otherwise, the instrument should be redesigned taking into account the balance between the reliability and the level of nuances we would like to obtain from observational data (Zink et al., 2007). According to Janouskova et al. (2022), the form and content have a role to play in how one processes the information received. Thus, when evaluating such complex skills like health communication in real-life settings or simulated scenarios with more than one participating actors, the format and content of evaluation should be simplified to facilitate the examiners/raters' cognitive processing. This, in turn, is more likely to result in objective assessment of nurse communication skills.

The test-retest reliability was not so evident in the sample. One possible explanation is that there is a lack of consistency over the course of four weeks (Kobak et al., 2004). Moreover, Zink et al. (2007) attribute the lack of test-retest reliability to the complexity of the items. The authors argue that achieving good intra-rater reliability is more difficult for items measured on a Likert scale than on a discrete scale, i.e., yes or no. This is because the former specifies a qualitative assessment rather than the presence of the trait or characteristics of interest. Thus, the low intrarater reliability may have been affected by the complexity of the items.

Regarding criterion validity, the result showed that the ratings of participants can be considered as being similar to those of experts in video 3 (low performance). Pagano et al. (2015) found that there were only three substantial differences between the two samples' ratings across three items by using Pearson correlations. By means of a Bland-Altman plot, we

could only support this finding to a lesser extent. This means that the overall mean difference between participants and experts' ratings was trivial only in low-performance demonstration of the communication skills, but not in the cases of high and medium skill exhibition.

### *Clinical Implications*

Attempting to validate the French-language version of the HCAT, our findings reveal that nurse educators can employ the instrument to evaluate nurse students' communication skills, considering the following recommendations. First, it can be established that the HCAT-f can capture three aspects of communication skills, namely professional presentation, empathy, and trust building. Nevertheless, due to its low single measure ICC, only the average scores can be used. In other words, interrater agreement at the factor level was found to be most consistent i.e., the average score from all the items constituting one of the three dimensions. Given that, an overall mean score of the instrument is best used as a summative assessment of students' communication. Examining individual items of the HCAT-f will provide formative assessment for students to identify specific areas they can work on to improve their competence in communication.

Second, our findings reveal that when it comes to high levels of skill demonstration, the raters tend not to strongly agree with one another. Thus, prior training as to which behaviours or gestures can be registered at a four or a five can be helpful to achieve raters' consensus.

Third, as communication skills encompassed empathy and trust besides professional presentation, we would suggest that the former is rather subjective, and context bound. Therefore, organising a briefing before the examination among the raters to obtain their consensus on the conceptualisation of these two constructs would help enhance the interrater agreement. Moreover, as a worldwide used tool, exploiting different language versions of the HCAT could help better understand the cultural differences encountered in clinical communication performance.

Indeed, clinical communication has become a key element of healthcare curriculum (Brouwers et al., 2019). In this sense, the lack of a common language or taxonomy for assessing communicative skills, as well as the inconsistent type and quality of reported results, heavily impairs a comparison of results of the scientific literature (Setyonugroho et al., 2015). Hence, it is nearly impossible to compare clinical communication performance results, identify potential associations or gaps, and therefore propose relevant interventions in the healthcare environment. While, on the one hand, the development of new communication assessment tools enriches the literature, on the other hand, it restricts comparisons with previous results. It seems appropriate in the future to identify standard tools validated in different languages to allow meaningful comparisons across the world.

### *Limitations*

The sample comes from a French-speaking community in Belgium; therefore, the result cannot be generalised. Further validation of the HCAT-f in another French-speaking context is needed, which will add to the applicability of the French version of the instrument across different cultures speaking the same language. Additionally, we suggest that further validation studies should focus on the conceptualisation of nurses' global communication skills and redesign of the instrument before large-scale validation. Due to availability, the experts' ratings were collected once, i.e., they provided the referential rating during wave 1 only. Thus, the

comparison between sample ratings and experts' ratings was only available for wave 1 dataset. That the experts and nurse educators rated three videos with different levels of skill performance has an inevitable memory effect. The length of the videos might have resulted in rating biases, particularly in the second round after the participants had been familiar with the three scenarios. Future research following this design can subset the samples, i.e., one group of participants rates only one of the three videos. In this way, the ratings can be less biased. Even if there is no particular reason that could bias the results over time, it should be noted that data collection was performed in 2018.

## Conclusion

The present study validated the HCAT-f on a sample of nurse educators in Belgium. The result shows that the instrument reflects a three-factor structure. This means that empathy, trust building, and professional presentation of nurses' communication skills can be captured. Across the three videos, higher intra-rater reliability and inter-rater agreement were found in cases of low and medium skills demonstration. However, the ICCs reflected only fair to good agreement. Our findings suggest that further works should be carried out at the theoretical and empirical levels. More specifically, a framework to operationalise nurses' communications skills viewed from a staged perspective can be helpful. Items developed to measure aspects of communication should balance the complexity and reliability so that users of the instrument can obtain adequate, valid, and reliable information for evaluation purposes. In the meantime, the HCAT-f can serve as a summative measurement of students' communication skills thanks to its excellent interrater reliability for average measures and item pools due to the established factor structure. In brief, by means of a validation study across different levels of communication skills demonstration, the HCAT-f can be a useful tool for its target users, i.e., nurse educators. In this respect, the training of nurses' communications in simulation settings in French-speaking countries can be enhanced by the availability of such a validated instrument given that such a tool has not existed prior to the present study.

## Acknowledgements

The study protocol was approved by the university of Liege's ethics committee, reference number 2017/256 dated October 17, 2017.

## Funding

The research received no public funding.

## Conflict of Interest

The authors have no conflicts of interests.

## References

- Alter, A. L., & Oppenheimer, D. M. (2009). Suppressing secrecy through metacognitive ease: Cognitive fluency encourages self-disclosure. *Psychological Science*, 20(11), 1414-1420. <https://doi.org/10.1111/j.1467-9280.2009.02461.x>
- Aranda, S., & Yates, P. (2009). Competency assessment tool for therapeutic communication. *The National Cancer Nursing Education Project (EdCaN)*, Cancer Australia. <http://edcan.org.au/assets/edcan/files/docs/EdCan-CAT-Therap-Comm.pdf>
- Baird, T., Corey, V. R., Sissom, J., Weber, J., & Campbell, S. H. (2021). Testing the validity and reliability of the adapted Health Communication Assessment Tool©. *Clinical Simulation in Nursing*, 52, 50-58. <https://doi.org/10.1016/j.ecns.2020.11.001>
- Bandalos, D. L., & Boehm-Kaufman, M. R. (2009). Common misconceptions in exploratory factor analysis. *Statistical and methodological myths and urban legends: Where pray tell did they get this idea*, 63-88.
- Baumann, M., Baumann, C., Le Bihan, E., & Chau, N. (2008). How patients perceive the therapeutic communications skills of their general practitioners, and how that perception affects adherence: use of the TCom-skill GP scale in a specific geographical area. *BMC Health Services Research*, 8(1), 244, 1-9. <https://doi.org/10.1186/1472-6963-8-244>
- Biggs, J. (2003, April). *Aligning teaching and assessing to course objectives*. Teaching and learning in higher education: New trends and innovations, University of Aveiro, Portugal.
- Bullinger, M., Anderson, R., Cella, D., & Aaronson, N. (1993). Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Quality of life Research*, 2(6), 451-459. <https://doi.org/10.1007/bf00422219>
- Byrne, B. M. (1998). *Structural equation modeling with Lisrel, Prelis, and Simplis*. Psychology Press.
- Campbell, S. H., Aredes, N. D. A., & Dhari, R. K. (2018). Teaching and evaluating therapeutic communication in simulated scenarios. In S. H. Campbell & K. Daley (Eds.), *Simulation scenarios for nursing educators: Making it real* (3rd ed., pp. 27-36). Springer Publishing Company.
- Campbell, S. H., Pagano, M. P., O'Shea, E. R., Connery, C., & Caron, C. (2013). Development of the health communication assessment tool: Enhancing relationships, empowerment, and power-sharing skills. *Clinical Simulation in Nursing*, 9(11), e543-e550. <https://doi.org/10.1016/j.ecns.2013.04.016>
- Caron, A., Perzynski, A., Thomas, C., Saade, J. Y., McFarlane, M., & Becker, J. (2013). Development of the objective, structured communication assessment of residents (OSCAR) tool for measuring communication skills with patients. *Journal of Graduate Medical Education*, 5(4), 570-575. <https://doi.org/10.4300/JGME-D-12-00230.1>
- Chen, X., Luo, L., Jiang, L., Shi, L., Yang, L., Zeng, Y., ... & Li, L. (2021). Development of the nurse's communication ability with angry patients scale and evaluation of its psychometric properties. *Journal of Advanced Nursing*, 77(6), 2700-2708. <https://doi.org/10.1111/jan.14788>
- Fleischer, S., Berg, A., Zimmermann, M., Wüste, K., & Behrens, J. (2009). Nurse-patient interaction and communication: A systematic literature review. *Journal of Public Health*, 17(5), 339-353. <https://doi.org/10.1007/s10389-008-0238-1>
- Foronda, C., MacWilliams, B., & McArthur, E. (2016). Interprofessional communication in healthcare: An integrative review. *Nurse Education in Practice*, 19, 36-40. <https://doi.org/10.1016/j.nepr.2016.04.005>

- Goes, F.S.N., Aredes, N.D.A, Fonseca, L.M.M., Hara, C.Y.N. & Campbell, S.H. (2017). Simulation with standardized patients: Nursing student health communication skills. *The Northeastern Nursing Network Journal*, 18(3), 383-389. <http://dx.doi.org/10.15253/2175-6783.2017000300014>
- Hargie, O., Dickson, D., Boohan, M., & Hughes, K. (1998). A survey of communication skills training in UK schools of medicine: Present practices and prospective proposals. *Medical Education*, 32(1), 25-34. <https://doi.org/10.1046/j.1365-2923.1998.00154.x>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Janouskova, A., Kocyan, J., Simova, M., Uvirova, K., Zahradnickova, K., Vaculik, M., & Prochazka, J. (2022). The effect of font readability on the Moses illusion: A replication study. *Consciousness and Cognition*, 99, Article 103284. <https://doi.org/10.1016/j.concog.2022.103284>
- Jung, S., & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior Research Methods*, 43(3), 701-709. <https://doi.org/10.3758/s13428-011-0077-9>
- Kardong-Edgren, S., Adamson, K. A., & Fitzgerald, C. (2010). A review of currently published evaluation instruments for human patient simulation. *Clinical Simulation in Nursing*, 6(1), e25-e35. <https://doi.org/10.1016/j.ecns.2009.08.004>
- Kerr, D., Martin, P., Furber, L., Winterburn, S., Milnes, S., Nielsen, A., & Strachan, P. (2022). Communication skills training for nurses: Is it time for a standardised nursing model? *Patient Education and Counseling*, 105(7), 1970-1975. <https://doi.org/10.1016/j.pec.2022.03.008>
- Kobak, K. A., Engelhardt, N., Williams, J. B., & Lipsitz, J. D. (2004). Rater training in multicenter clinical trials: Issues and recommendations. *Journal of Clinical Psychopharmacology*, 24(2), 113-117. <https://doi.org/10.1097/01.jcp.0000116651.91923.54>
- Mendi, O., Yildirim, N., & Mendi, B. (2020). Cross-cultural adaptation, reliability, and validity of the Turkish version of the health professionals communication skills scale. *Asian Nursing Research*, 14(5), 312-319. <https://doi.org/10.1016/j.anr.2020.09.003>
- O'Halloran, R., Grohn, B., & Worrall, L. (2012). Environmental factors that influence communication for patients with a communication disability in acute hospital stroke units: A qualitative metasynthesis. *Archives of Physical Medicine and Rehabilitation*, 93(1), S77-S85. <https://doi.org/10.1016/j.apmr.2011.06.039>
- Pagano, M. P., O'Shea, E. R., Campbell, S. H., Currie, L. M., Chamberlin, E., & Pates, C. A. (2015). Validating the Health Communication Assessment Tool©(HCAAT). *Clinical Simulation in Nursing*, 11(9), 402-410. <https://doi.org/10.1016/j.ecns.2015.06.001>
- Pallant, J. (2010)(4<sup>th</sup> Ed). *Survival Manual: A step by step guide to data analysis using SPSS program*. Open University Press.
- 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. <https://doi.org/10.1002/nur.20199>

- Reis, N. B. C., Góes, F. S. N., Aredes, N. D. A., Campbell, S. H. (2018). Adaptação cultural da ferramenta de avaliação de comunicação em saúde (HCAT) para a língua portuguesa, Brasil. [Cultural adaptation of the Health Communication Assessment Tool (HCAT) to Portuguese language, Brasil.] *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, 12, 444-455. <https://doi.org/10.29397/reciis.v12i4.1501>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 1904-1920, 3(3), 271-295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., ... & de Vet, H. C. (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>
- Vallerand, R. J. (1989). Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue française. *Canadian Psychology/Psychologie Canadienne*, 30(4), 662-680. <https://doi.org/10.1037/h0079856>
- Williams, C. A., & Gossett, M. T. (2001). Nursing communication: Advocacy for the patient or physician? *Clinical Nursing Research*, 10(3), 332-340. <https://doi.org/10.1177/c10n3r8>
- Yang, H. M., & Hwang, S. Y. (2016). Reliability and validity of the assessment tool for measuring communication skills in nursing simulation education. *Korean Journal of Adult Nursing*, 28(1), 95-105. <https://doi.org/110.7469/kjan.2016.28.1.95>
- Zenk, S. N., Schulz, A. J., Mentz, G., House, J. S., Gravlee, C. C., Miranda, P. Y., ... & Kannan, S. (2007). Inter-rater and test-retest reliability: Methods and results for the neighborhood observational checklist. *Health & Place*, 13(2), 452-465. <https://doi.org/10.1016/j.healthplace.2006.05.003>

## Author Contributions

Conceptualisation (main idea, theory): Vanessa Wauthier, Méryl Paquay, Jean-Christophe Servotte, Jonathan Goffoy, & Alexandre Ghuysen

Funding acquisition: Not applicable

Project administration: Jean-Christophe Servotte, Anne-Françoise Donneau, & Alexandre Ghuysen

Methodology (design, operationalisation): Vanessa Wauthier, Méryl Paquay, Jean-Christophe Servotte, Sophie Vanbelle, Suzanne Hetzel Campbell, Jonathan Goffoy, & Alexandre Ghuysen,

Data collection: Vanessa Wauthier, Méryl Paquay, & Jean-Christophe Servotte

Data analysis: Anh Nguyet Diep & Nadia Dardenne

Writing – original draft: Anh Nguyet Diep, Jean-Christophe Servotte & Anne-Françoise Donneau

Review & editing: all authors

## Author biographies

**Anh Nguyet Diep** works as a biostatistician at the Biostats Unit, Department of Public Health, University of Liège. Her research focuses on blended learning with students' learning and satisfaction, online interaction and instrument validation as major topics.

**Jean-Christophe Servotte** 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 and lecturer at the Paramedical Department, Namur-Liège-Luxembourg University College. He is Chair of the INACSL International Committee.

**Nadia Dardenne** 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.

**Sophie Vanbelle** is professor at the department of Methodology & Statistics, School for Public Health and Primary Care, Faculty of Health, Medicine and Life Sciences, Maastricht university. Her research focuses on multi-agreement and longitudinal modelling as the main topics.

**Vanessa Wauthier** obtained her master's degree in Public Health, University of Liège. She is currently working at Centre Hospitalier Régional de Namur as a coordinator of the cardiac catheterisation room.

**Méryl Paquay** is a nurse with a specialisation in community health and a master's degree in public health. She is also certified to work as a simulation instructor. She has recently started a PhD. Her area of expertise covers topics related to Crisis Resources Management, reporting, simulation and the impact of the work environment on the quality of care.

**Suzanne Hetzel Campbell** is professor at the School of Nursing, the University of British Columbia. Her research focuses on innovative teaching methodologies and her increasing interest in web-enhanced learning and simulation-based pedagogy has led to research, publications and workshops on these topics empowering nursing faculty, students, as well as healthcare professionals.

**Jonathan Goffoy** is perfusionist at CHU de Liège. He works also as simulation instructor and research assistant at the Simulation Center, University of Liege.

**Anne-Françoise Donneau** is full professor of biostatistics at the Faculty of Medicine at the University of Liège and Director of the Biostatistics and Research methods Unit. Her research interests include ordinal variables, missing data, multiple imputation, longitudinal/correlated data, cluster methods and validation techniques.

**Alexandre Ghuysen** is 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.