Brazilian version of the Calgary-Cambridge Observational Guide (CCOG): cross-cultural adaptation and psychometric properties

MARCELA DOHMS
Universidade de Sao Paulo

Carlos F Collares
Universiteit Maastricht

Iolanda Tibério (iocalvo@uol.com.br)
Universidade de Sao Paulo https://orcid.org/0000-0002-5662-7895

Research article

Keywords: Communication skills, Assessment, Medical Education, Psychometrics

Posted Date: June 28th, 2019

DOI: https://doi.org/10.21203/rs.2.10774/v1

Abstract

Introduction The search for appropriate tools to assess communicational skills remains an ongoing challenge. The Calgary-Cambridge Observational Guide is designed for measuring and comparing performance in communication skills training Objective Adapt the 28-item version of the guide for the Brazilian cultural context and perform a psychometric quality analysis of the tool. Method Experienced preceptors (35) evaluated videos of 5 medical residents with a simulated patient. using the translated version. For the cultural adaptation we followed the methodological norms on synthesis, retro-translation, committee review, and testing. We conducted the data analysis with the Rasch Many-Facet Model. Results The reliability of the internal consistency was satisfactory, with the Cronbach's alpha coefficient acceptable in all 5 evaluations (0.88; 0.84; 0.89; 0.87; 0.83). The reliability coefficient was high in the evaluators' facets (0.90), the stations (0.99), and items (0.98). The data indicates the suitability of the tool to the Rasch Many-Facet Model. The evaluators had greater difficulty with attitudinal items, such as demonstration of respect, confidence, and empathy. Conclusion The psychometric properties of the tool were adequate. The reliability indicators of the Rasch Model presented a good potential for reproducing the Brazilian version of the tool as well as acceptable reliability for its use.

Background

Communication skills are essential components in medical training, and one of our present challenges lies in how to evaluate these skills in medical training and finding the most appropriate tools. The vast majority of the most commonly used tools for evaluating communication skills have been developed as a checklist for observation and usage in summative evaluations in OSCE-type examinations for performance comparisons and formative feedback[1].

The Calgary-Cambridge Observational Guide (CCOG) is a certified questionnaire created to measure and compare performance in communication skills training. First published in 1996 in Canada, the guide served for the curriculum planning when teaching communication skills[2]. The complete observational guide (CCOG) has 71 items which makes its application considerably difficult. Therefore, we opted for its summarized version, with 28 items, developed for evaluating OSCE-type stations. This abridged version presented adequate psychometric properties in previous studies. The 28-item CCOG questionnaire is divided into 6 blocks/domains, according to the stages of the query with a 3-point scale ("YES", "YES, BUT" and "NO") in a checklist format[1].

Objective

Our study sought a translation and cross-cultural adaptation to Brazilian Portuguese of the abridged 28-item version of the CCOG questionnaire for use in the evaluation of communication skills in medical training as well as to analyze the psychometric quality of the tool and present the initial validation results.

Methods

The Ethics Committee of the Municipal Health Secretariat of Rio de Janeiro approved this research project (CAAE: 57387816.7.0000.5279). A Free and Informed Consent Term with explanations about the research was presented and filled out by the research subjects at this stage.

The cultural adaptation followed the methodological norms recommended by researches by way of the following stages: translation; synthesis; retro–translation; review by a committee; and pre-testing. The first stage was the translation of the tool from English to Brazilian Portuguese. To this end, two bilingual translators with native Portuguese performed two independent translations. After the translations, the tool was sent to a Committee of Experts of 6 physicians with teaching experience in communication skills to validate the content and differences in both translations. Each committee member answered the questionnaire on each translated item by means of the online platform googleform and contributed with suggestions. Afterwards, we merged all versions of the tool and indicated which features should be present in the pre-test version. After reaching a consensus, we took into account the semantic, idiomatic, experimental, and conceptual equivalences. After all suggestions, a pre-test version was developed. The tool was then applied to the research subjects, 35 preceptors in a Primary Care program.

Initially, all participants received instructions on how to fill out the questionnaire and presented their questions on all items of the questionnaire. Subsequently, they all watched at the same time the 5 interview videos of 5 doctor-residents from a same OSCE station with a simulated patient. Each video had a maximum running time of 7 minutes. Then they answered the 28-item CCOG questionnaire for each video. After completing and submitting the questionnaires a debate took place regarding the understanding of each item and the difficulties encountered when answering. Suggestions for altering any item were duly noted and later discussed among the authors for modifications in the final version.

To evaluate the inter-rater reliability, for each domain of the questionnaire we estimated the intra-class correlation coefficient (ICC). To evaluate the internal consistency of each domain, we estimated the Cronbach's alpha coefficient. Within each domain, in order to evaluate the degree of importance of each question, we estimated the Cronbach's alpha coefficient with the exclusion of each question. We evaluated the correlation between the domains of the questionnaire by estimating the Spearman correlation coefficient and testing its significance. Values of $p < 0.05$ indicated statistical significance. We then analyzed the data by using the IBM SPSS Statistics v.20 software.

Under weighted-least squares mean- and variance-adjusted (WLSMV) estimation, categorical confirmatory factor analyses analogous to the Samejima graded response model were performed using unidimensional and multidimensional models for the comparison of goodness-of-fit indices, namely chi-square, root mean square error of approximation (RMSEA), the confirmatory fit index (CFI), the Tucker-Lewis index (TLI) and the weighted root mean square residual, besides a chi-square difference test. The chi-square/df ratio was considered good when below 2,0 and acceptable when below 2,5[3]. RMSEA values were considered satisfactory if below 0.10 and good if below 0.08[4]. Both CFI and TLI values were

considered satisfactory if above 0.90[5]. WRMR was considered satisfactory if below 1.0 (23). Mplus 8.0 was used for this analysis.

To provide additional evidence to support the validation of the interpretation of the scores, we also analyzed the data by using the Many-Facet Rasch Model (MFRM) developed by Linacre[6]. In this analytical model for measuring the psychometric quality of the tool, we are able to analyze how multiple variables may have a simultaneous influence on the scores, allowing us to estimate a completely neutral examiner and provide an estimated “fair score”[7]. The MFRM model has been increasingly used to analyze the quality of evaluations with response items as it allows for the inclusion of other important variables that may be bias generators in evaluation processes, such as the personal characteristics of the evaluators, their propensities, and criteria differences regarding severity or understanding[8]. The model attempts to calibrate items regardless of the persons involved[9]. The data was analyzed with the FACETS software, version 3.71.4.

Results

Descriptive statistics, Cronbach's alpha coefficients, and intra-class correlation coefficient

We observed (table 1) that the item with the highest agreement among evaluators was item 23 (Demonstrates no prejudice or judgment) and the ones with the least agreements were item 4 (Identifies and confirms problems list), 12 (Establishes dates and sequence of events), and 28 (Contracts with patient the next steps). In many items we find a similar percentage between “yes” and “yes, but...”.

The Cronbach alpha coefficient analyzes the homogeneity of the items by measuring the reliability of the questionnaire based on the internal consistency of its items. The value ranges from 0 to 1, with zero being the total absence of internal consistency and 1 being perfect consistency. A coefficient above 0.7 is considered acceptable[10]. We found that the index was high for all video (table 2). We also estimated the Cronbach alpha coefficient for each block of questions (domains), considering all questions in the block as well as the exclusion of each question from the block.

In the analysis of the Cronbach's alpha coefficients and considering all questions in the block as well as considering the exclusion of each question from the block, we observed that the questions that mostly increased the reliability if taken from at least 2 videos were: 1 (Greets patient), 8 (Listens attentively), 11 (Clarifies patient’s statements), 16 (Perceives and reacts to verbal and non-verbal clues), 20 (Uses time efficiently.), 22 (If reads or writes, does so without interfering with dialogue/rapport), and 26 (Encourages patient to discuss additional issues).

The domains with the best ICC (table 3) were 2 (Exploring problems) and 3 (Evaluation of the patient's representations) while the worst was 6 (Concluding the consultation).

Correlation analysis between domains

Another important coefficient for validating a questionnaire is the correlation coefficient (r), which evaluates the intensity of the relationship between domains. The higher and closer means that observations have more similar classifications between the 2 variables. We conducted an analysis considering the sum of points for each domain in each station (table 4).

Confirmatory factor analysis and Many-Facet Rasch Mode analysis.

The confirmatory factor analysis under item response theory showed a favorable comparison to the proposed theoretical model in comparison to a unidimensional model. The χ^2/df ratio was 1.92 for the theoretical model and 2.60 for the unidimensional model, suggesting that an adequate fit was obtained by the proposed model and not for the latter[3]. The CFI was 0.90 for the theoretical model, thus above the acceptable threshold[11] and 0.82 for the unidimensional model, a result that reinforces the validity of the tool based on its internal structure. TLI (Tucker-Lewis Index) was 0.89 for the proposed model and 0.81 for the unidimensional model[4]. The RMSEA (Root Mean Square Error of Approximation) for the proposed model was 0.068, with a 90% confidence interval between 0.060 and 0.076, thus within the desirable threshold, which is up to 0.08[4]. The unidimensional model had an RMSEA of 0.089 with a 90% confidence interval between 0.082 and 0.096. The WRMR index, which is the weighted residual mean square root, was 1.21, above the desirable limit of 1.0[11], but more unfavorable to the unidimensional model, which had a WRMR of 1.56. A chi-square difference testing was used to assess whether the proposed multidimensional model had a significantly better fit than the unidimensional model. The test showed a significant better fit for the model proposed by the tool (χ^2 (15) = 210.510, $p < 0.001$). The loadings were -0.003 and -0.004 for items 1 and 2 respectively, revealing they were non-informative for the composition of scores. All the other items had discriminations above 0.3. The exclusion of these items improved fit very marginally, with CFI raising to 0.91 and TLI raising to 0.90 in the theoretical model proposed. A bifactor model was attempted, given the degree of correlation observed between the latent variables but the sample size did not allow convergence. Graph 1 contains the diagram with the multidimensional factor structure of the CCOG.

The MFRM model provided good evidence of the validity of the CCOG based on its internal structure. Unidimensional Rasch measures explained 30.88% of the score variance. Residual correlations were below the tolerable limit of 0.7, thus making the CCOG with an acceptable degree of local independence. However, principal component analysis of the MFRM residuals revealed two residual components that explained more than 5% of the variance (5.8 and 5.2%, respectively), a result convergent to the

multidimensionality observed in the confirmatory factor analysis. However, one could argue in favor of some degree of essential unidimensionality, given the little amount of variance explained by the first two residual components. The usefulness of applying the procedure regardless of the violation of the absolute unidimensionality assumption consists in using the probabilistic approach of the Rasch model to place more facets of the assessment tool in the same latent scale. The examinee's ability increases the probability of endorsement in the scale and the other elements are estimated to decrease the probability of endorsement. On Graph 2, it is possible to observe that Station 2 was the most difficult and Station 3 was the easiest. The graph also demonstrates a good overlap between examinee's abilities, stations and items, which has likely contributed to the positive results observed in terms of reliability.

MFRM analysis enables the estimation of separate reliability estimates for each facet of the assessment with inflation for misfit. Reliability for examinees was 0.90. Individual reliability estimates calculated using the standard error of measurement of MFRM latent scores revealed that except for one examinee, all participants had score reliabilities above 0.80 as it can be seen on Graph 3. MFRM-based reliability for stations was 0.99 and, for items, 0.98. Fit indices for examinees, stations and items was also satisfactory. All examinees, items and stations had *infit* and *outfit* indices between the optimal range of 0.5 and 1.5, with the exception of one examinee, who had an outfit value of 1.68. None of the fit indices were above 2.0, a point in which one can expect a degraded quality of the measures.

Discussion

The analysis of the cultural adaptation process of the internationally validated CCOG questionnaire showed good results. The Cronbach alpha value for all videos taken together was above 0.80, following the preference established by Streiner[10], who suggests that coefficient values should be between 0.80 and 0.90, indicating that the Brazilian version shows acceptable internal consistency and reliability.

Items 1 (Greets patient) and 2 (Introduces self and role) of the tool showed negative loads in the confirmatory factor analysis, which suggests that they are not adequately measuring the intended construct. The analysis shows that 0% of the variance of these two items may be explained by the intended construct as 100% noise or other non-intended construct. The determination coefficient (R^2), which tells us how much variance of a variable is explained by another variable (in our case the item verse construct) also had a low value (zero). This may be explained by the fact that it was a simulated station and the fact that the recording began inside the office, but some resident doctors greeted the patient before entering the office and began filming, thus interfering with the analysis. Since the evaluators were instructed to leave blank if the task could not be evaluated, these initial items for starting the consultation had the most amount of blanks.

When redoing the same type of analysis without items 1 and 2, we see an improvement in the model adjustment. The adjustment indices of the observed data in relation to the proposed theoretical model,

together with the significant improvement of the adjustment obtained with the model in comparison to a one-dimensional model are validity evidences based on the internal structure of the tool. Thus, we suggest that in order to evaluate items 1 and 2, the beginning of the consultation should take place inside the evaluated environment so that we may observe the interviewer greeting and introducing themselves to the patient.

The item with the highest agreement among evaluators was that resident did not show judgment, which seems to be a clearer parameter and residents are well trained to avoid. Among the items with the highest disagreement among respondents is an item with 2 tasks: “identify and confirm the list of issues”. Perhaps having two tasks on the same item interfered with the difference in responses. For this reason, we suggest changing the item to “confirm the list of problems”, since to confirm the problems the student or resident must have already previously identified them.

We found significant disagreements in the task “Establish dates”, which the reviewers suggested changes. Most likely, the difference in answers occurred due to the difficulty in understanding and so we modified the final version. The last item of the questionnaire, which refers to the important ability of making a shared decision, also showed a high degree of disagreement, probably because a complete agreement with the patient involves a complexity of dialogues and negotiations that may need better defined parameters. There was low intra-class correlation coefficient in the domain “Closes the consultation”, probably because of the difficulty in understanding the shared decision-making process in the item “Contracts with the patient the next steps”. The word “Contract” may give room for different interpretations as to what one considers a satisfactory degree of patient participation in the decision-making.

Other items have a more subjective interpretation and caused greater differences in the evaluation. The evaluators mentioned difficulties when defining parameters in less objective or technical behavioral evaluation items, such as “Demonstrates respect / appears confident / demonstrates empathy”. This may be because such items need further development when defining their parameters among evaluators, according to the learning objectives for each phase of medical training. Moreover, because they are complex tasks and difficult to judge as an external observer, who attempts to measure the feelings of the interviewer and the patient. For a wholly real evaluation we would need to know the patient's opinion, such as if the patient felt confidence in the interviewer or felt that the interviewer was empathetic.

These items in particular should be discussed among the group of evaluators to define what will be considered satisfactory or unsatisfactory and partially satisfactory. We observed that when an item was not performed, the evaluators found it easier to evaluate as "No", but when the residents performed the task, the evaluators were in doubt about choosing between “YES” and “YES, BUT...”, thus indicating that we need to better define when a task is accomplished wholly or partially.

These difficulties may have interfered in some reliability and validity coefficients of the tool. However, another scale-validation study in Germany showed similar intra-class correlation coefficients ranging from 0.05 to 0.57. In this study, the authors suggested deleting the item “Negotiates agenda” when using

the questionnaire with students at the beginning of the course and only using it toward the end of the course, when they would have the ability to address multiple topics and perform other procedures besides merely collecting the patient's history. In addition, they also attributed the reliability difficulties of the scale in the study due to the need for further instructions and better defined parameters among evaluators before the application[12].

In addition, the difficulties presented when evaluating and judging the items may hamper a final summative evaluation. Increasing discussions have taken place regarding the association of judging by items with a subjective holistic judgment[13]. Some studies comparing the psychometric properties of checklists and global assessment scales in OSCEs evaluated by experts indicated a higher reliability between stations and a better validity than checklists[14,15]. The tool predicts, in its original version, a global evaluation with no note value between “SATISFACTORY”, “SATISFACTORY, BUT...” and “UNSATISFACTORY”, which we did not use in the study, but recommend its regular use alongside the questionnaire.

The tool may be used for both a formative and summative evaluation, but due to the difficulties already discussed in the judgment of more subjective items, CGOC may bring about more benefits when applied to a formative rather than summative evaluation. We should emphasize that, whenever possible, constructive and detailed narrative feedback should be associated with a summative assessment, since students prefer feedback rather than notes and the effectiveness of reflective feedbacks is high[16].

Despite the difficulties observed, the reliability coefficients in the Many-Facet Rasch Model were excellent in all facets, varying from 0 to 1 and the higher the value, the lower the risk of false positives or negatives and thus a lower measurement error, demonstrating that the tool has acceptable reliability for reproducibility in other contexts. One limitation of this study is that evaluators, while all experienced preceptors, had different backgrounds in assessing communication skills. In addition, although we had about 170 tools filled, a larger sample could provide more information. Another limitation is that we were unable to conduct a second evaluation to confirm the reliability of the evaluation among each evaluator. In addition, the preceptors’ evaluation could have been associated and/or compared with the evaluation from other standpoints, such as colleagues, staff, and patients, since evaluation by multiple sources and at different times in the Medical Residency Programs have shown to be good evaluators of attitudinal skills and complex tasks[17,18].

Although the study consisted of resident physicians, we believe the tool can also be used with undergraduate students, as previously demonstrated in other studies, with adaptations and standardization of the parameters of items according to the course period and learning objectives[12]. We underline the extreme importance of discussing with the group of evaluators each word in the questionnaire and its subsequent use in practice for constant improvements, which should continue to suffer adjustments with the feedback. We suggest further researches on evaluation tools for attitudinal skills, enhancing the definition of subjective items according to the learning objectives of each medical training phase.

As the validity of a tool is a continuous process[19,20], and thus we recommend that scale items be continually reevaluated for constant improvements, and we emphasize the importance of homogenizing evaluation parameters among evaluators on each item prior to scale application, clarifying the learning objectives required for each training level, especially when it comes to less objective attitudinal evaluation items, such as demonstration of respect, confidence and empathy. In addition, we also suggest complementing the evaluation of communication skills by considering other sources and viewpoints such as patients, colleagues, and staff.

Conclusions

The Brazilian version of the summarized CCOG used in this study has demonstrated evidence of reliability and validity based on internal structure for evaluating communication skills in Brazil. We had limitations regarding to evaluators and sample size. We suggest further instructions and better-defined parameters among evaluators before the application and to complement the assessment with an overall evaluation with a subjective holistic judgment. Since the consequential validity seems to be greater in the formative evaluation[21], we also suggest associating the scale with a detailed narrative feedback and a continual reevaluation for constant improvements.

Abbreviations

CCOG - Questionnaire based in Calgary-Cambridge Observation Guide

ICC - intra-class correlation coefficient

MFRM -Many-Facet Rasch Model

(r) - Spearman correlation coefficient

CFI - comparative adjustment index

TLI -Tucker-Lewis Index

RMSEA - Root mean square error of approximation

WRMR - Residual mean square root

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Municipal Health Secretariat of Rio de Janeiro approved this research project (CAAE: 57387816.7.0000.5279). A written consent for participation was obtained from all participants involved in the study. The participation was voluntary.

Consent for publication

A written consent for publication was obtained from all participant**s** involved in the study.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Authors' contributions

MCD conceived the study, participated in data collection, performed statistical analysis, and drafted and revised the manuscript. CFC conceived the study, performed statistical analysis, and revised the manuscript. ICT conceived the study, participated in data collection, and revised the manuscript. All authors read and approved the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Kurtz S, Silverman J, Draper J. Teaching and learning communication skills in medicine. 2nd Oxford: Radcliff Publishing; 2005.
2. Silverman JD, Kurtz SM, Draper J. The Calgary-Cambridge approach to communication skills teaching. 1. Agenda led outcome-based analysis of the consultation. Educ Gen Pract. 1996; 7:288-

99.

3. Ullman J. B. Structural equation modeling, in*Using Multivariate Statistics*, eds Tabachnick B. G., Fidell L. S., editors. (Boston, MA: Allyn & Bacon); 2001. 653–771.
4. Browne MW, Cudek R. Alternate ways of assessing model fit. In Bollen K. A.; Long J. S. (Eds.). Testing structural equation models. Newbury Park, CA: Sage; 1993.
5. *Bentler*, P. M., *Bonett*, D. G. Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin *1980*, 88:588-606.
6. Linacre JM. Facets computer program for many-facet Rasch measurement, version 3.80.0. Beaverton, Oregon: Winsteps.com. Books, 2017.
7. Bond T, Fox C. Applying the Rasch Model Fundamental Measurement in the Human Sciences (2nd Edition), New York & London: Routledge; 2012.
8. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press; 1980.
9. Engelhard G. Evaluating the bookmark judgments of standard-setting panelists. J Educ and Psychol Measure. 2011;71:909-924.
10. Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn´t matter. Journal of Personality Assessment. 2003; 80:217-222,
11. Hu, LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal. 1999;6:1-55.
12. Simmenroth-Nayda A, Heinemann S, Nolte C, Fischer T, Himmel W. Psychometric properties of the Calgary Cambridge guides to assess communication skills of undergraduate medical students. International Journal of Medical Education. 2014; 5: 212-218.
13. Inayah AT, Anwer LA, Shareef MA et al. Objectivity in subjectivity: do students' self and peer assessments correlate with examiners' subjective and objective assessment in clinical skills? A prospective study. BMJ Open. 2017; doi: 10.1136/bmjopen-2016-012289.
14. Hodges B, Regehr G, Hanson M, Mcnaughton N. An objective structured clinical examination for evaluating psychiatric clinical clerks. Acad Med. 1997; 72:715-21.
15. Hodges B, Regehr G, Mcnaughton N, Tiberius R, Hanson M. Checklists do not capture increasing levels of expertise. Acad Med. 1999;74:1129-1134.
16. van der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB. Twelve tips for programmatic assessment. Med Teach. 2014; 37:641-646.
17. Moonen-Van Loon JM, Overeem K, Govaerts MJ, Verhoeven BH, van der Vleuten CP, Driessen EW. The reliability of multisource feedback in competency-based assessment programs: the effects of multiple occasions and assessor groups. Acad Med. 2015;90:1093-9.
18. Ten Cate O, Sargeant J. Multisource feedback for residents: how high must the stakes be? J Grad Med Educ. 2011;4:453–5.

19. Shepard LA. Chapter 9: Evaluating Test Validity. Review of Research in Education. 19. Washington, DC: AERA I; 1993. p. 405 – 450.
20. 20- Streiner DL, Geoffrey RN, Cairney J. Health Measurement Scales: A practical guide to their development and use paperback. Oxford University Press, Fifth Edition; 2014.
21. Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77:81–112.

Tables

Table 1 – Descriptive statistics of responses to the CCOG questionnaire with percentages according to response options.

		Analyzed video				
Questionnaire Item	**Evaluators' response**	**1**	**2**	**3**	**4**	**5**
1. Greets patient.	No	0	76	74	6	18
	Yes, but...	17	21	7	9	12
	Yes	83	31	19	85	69
2. Introduces self and role.	No	83	94	96	94	100
	Yes, but...	11	6	4	6	0
	Yes	6	0	0	0	0
3. Demonstrates respect.	No	0	13	3	0	6
	Yes, but...	57	77	6	23	42
	Yes	43	9	91	77	52
4. Identifies and confirms problems list.	No	27	50	21	40	22
	Yes, but...	48	37	29	40	53
	Yes	24	13	50	20	25
5. Negotiates agenda (reasons for consultation).	No	57	92	48	75	63
	Yes, but...	36	8	24	21	18
	Yes	7	0	28	4	18
6. Encourages the patient to tell story.	No	26	42	0	9	10
	Yes, but...	47	37	24	29	48
	Yes	27	21	76	62	42
7. Appropriately moves from open to closed questions.	No	53	76	15	53	42
	Yes, but...	30	15	37	29	52
	Yes	17	9	47	18	6
8. Listens attentively.	No	20	41	3	3	9
	Yes, but...	48	50	12	44	56
	Yes	31	9	85	53	34
9. Facilitates patient’s responses verbally and non-verbally.	No	35	83	0	38	36
	Yes, but...	35	17	33	38	61
	Yes	29	0	67	24	3
10. Uses easily understood questions and comments.	No	3	12	0	9	10
	Yes, but...	21	39	9	61	50
	Yes	76	49	91	30	40
11. Clarifies patient's statements.	No	10	22	7	39	22
	Yes, but...	45	61	36	50	55
	Yes	45	16	57	11	22
12. Establishes dates and sequence of events.	No	43	50	17	37	21
	Yes, but...	29	32	27	33	27
	Yes	28	18	56	30	52
13. Determines and acknowledges the patient's ideas regarding cause.	No	23	12	12	52	10
	Yes, but...	17	72	24	33	51
	Yes	60	16	64	15	39
14. Explores the patient's concerns about the problem.	No	6	13	15	47	12
	Yes, but...	17	56	21	37	41
	Yes	77	31	64	16	47
15. Encourages the patient to verbalize how they feel.	No	42	65	28	81	52
	Yes, but...	23	31	28	19	28
	Yes	35	4	44	0	20
16. Perceives and reacts to verbal and non-verbal clues.	No	43	78	42	79	42
	Yes, but...	48	22	35	21	42
	Yes	9	0	23	0	16
17. Summarizes at end of a specific line of inquiry	No	32	62	31	50	39
	Yes, but...	36	31	31	28	42
	Yes	32	7	38	22	19
18. Progresses using transition phrases between topics.	No	46	68	38	70	58
	Yes, but...	36	28	23	23	32
	Yes	18	28	39	7	10
19. Structures a logical sequence.	No	9	29	21	73	27
	Yes, but...	32	32	21	12	52
	Yes	59	39	58	15	21

20. Uses time efficiently.	No	0	21	6	61	41
	Yes, but...	20	46	23	29	24
	Yes	80	32	71	10	35
21. Demonstrates appropriate non-verbal behavior.	No	38	70	3	38	35
	Yes, but...	32	18	18	31	55
	Yes	55	12	79	31	10
22. If reads or writes, does so without interfering with dialogue/rapport.	No	8	80	4	70	58
	Yes, but...	23	1	11	18	24
	Yes	69	7	85	12	18
23. Demonstrates no prejudice or judgment.	No	11	4	0	11	19
	Yes, but...	21	29	11	26	23
	Yes	68	67	89	63	58
24. Displays empathy and support for the patient.	No	3	24	6	17	16
	Yes, but...	43	65	45	62	68
	Yes	54	10	49	21	16
25. Appears confident.	No	3	25	0	49	28
	Yes, but...	38	53	15	27	53
	Yes	59	22	85	24	19
26. Encourages patient to discuss additional issues.	No	40	50	78	33	52
	Yes, but...	24	28	19	37	29
	Yes	36	22	3	30	19
27. Closes consultation with a brief summary.	No	47	63	35	24	27
	Yes, but...	27	25	24	46	30
	Yes	26	12	41	30	43
28. Contracts with patient the next steps.	No	41	55	12	23	16
	Yes, but...	28	35	36	42	34
	Yes	31	10	52	35	50

Table 2 – General Cronbach's Alpha Indexes for each analyzed video.

Analyzed Video	General Cronbach's Alpha
Resident 1	0,88
Resident 2	0,84
Resident 3	0,89
Resident 4	0,87
Resident 5	0,83

Table 3 – Intra-class correlation coefficients (ICC) for each domain (in percentage).

Domain	Theme	ICC
1	Beginning the consultation.	0,36
2	Exploring problems.	0,45
3	Understanding the patient’s perspective.	0,27
4	Structuring the consultation.	0,32
5	Building the relationship.	0,45
6	Concluding the consultation.	0,06

Table 4 - Result of the domains across different Stations.

Domain	Station 1		Station 2		Station 3		Station 4		Station 5	
	R	P value	R	P value	r	P value	R	P value	r	P value
d1 x d2	0,68	0,001	0,03	0,900	0,44	0,128	0,59	0,017	0,64	0,014
d1 x d3	0,38	0,131	-0,06	0,828	0,39	0,100	0,41	0,085	0,15	0,567
d1 x d4	0,35	0,142	0,24	0,345	0,32	0,210	0,77	0,000	0,39	0,115
d1 x d5	0,46	0,063	0,15	0,558	0,06	0,813	0,53	0,019	0,60	0,011
d1 x d6	0,27	0,186	0,00	0,995	0,09	0,700	0,24	0,280	0,08	0,735
d2 x d3	0,52	0,072	0,40	0,092	0,39	0,135	0,29	0,280	0,78	0,001
d2 x d4	0,53	0,033	0,39	0,107	0,33	0,248	0,55	0,022	0,37	0,140
d2 x d5	0,69	0,004	0,78	0,000	0,07	0,777	0,81	0,000	0,36	0,186
d2 x d6	0,50	0,020	-0,20	0,368	0,48	0,042	0,22	0,356	-0,13	0,614
d3 x d4	0,65	0,005	0,38	0,168	0,22	0,416	0,71	0,001	0,43	0,065
d3 x d5	0,30	0,335	0,78	0,000	0,46	0,070	0,22	0,370	0,40	0,100
d3 x d6	0,12	0,640	0,14	0,555	0,12	0,616	0,03	0,907	0,00	0,995
d4 x d5	0,15	0,579	0,45	0,079	0,36	0,169	0,43	0,072	0,32	0,154
d4 x d6	-0,14	0,533	0,48	0,028	0,55	0,012	0,23	0,299	0,42	0,041
d5 x d6	0,25	0,323	-0,10	0,660	0,46	0,042	0,06	0,792	0,07	0,769

r: Spearman correlation coefficient

Figures

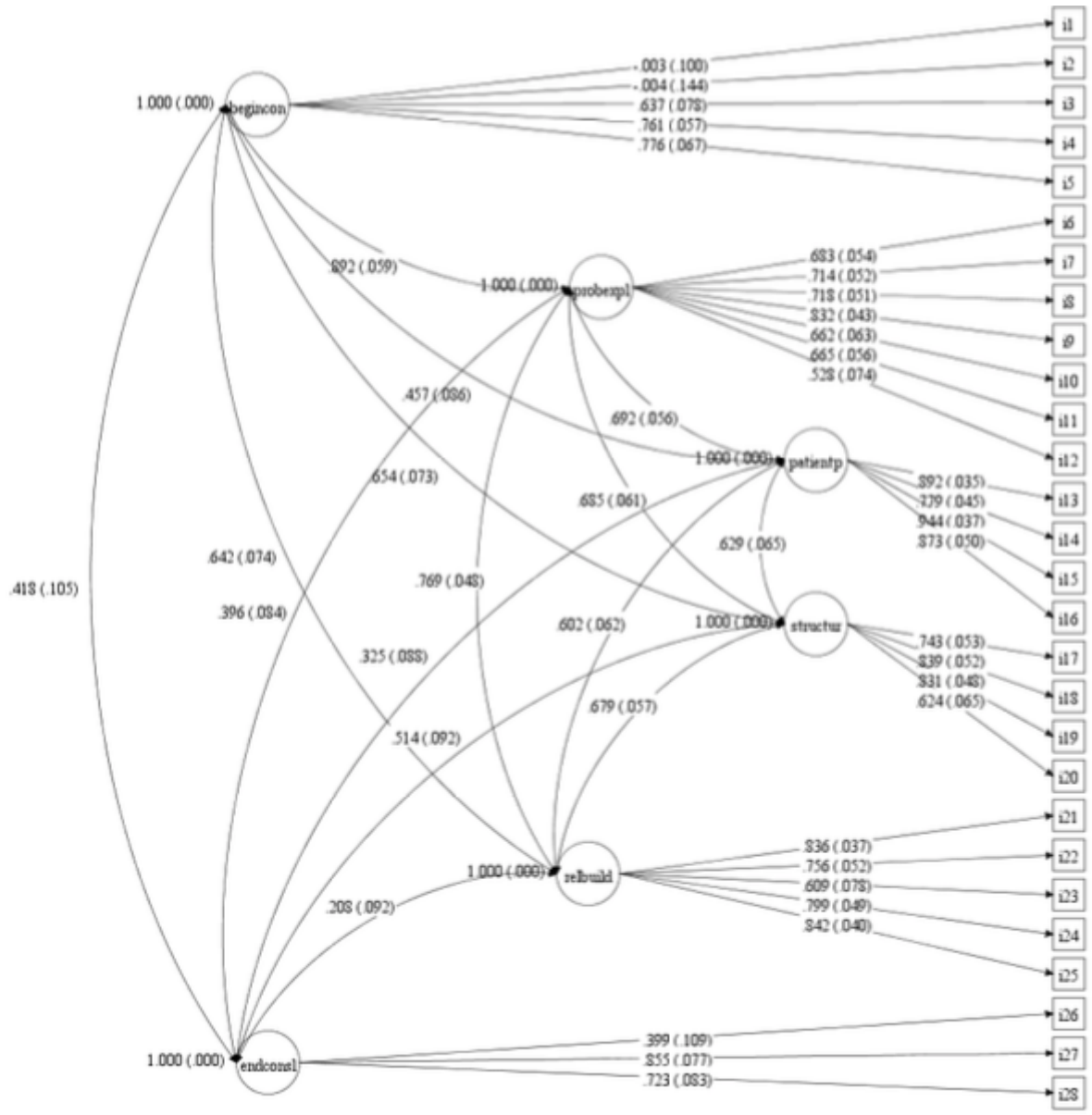

Figure 1

Graph 1. Diagram with the multidimensional item response theory-based confirmatory factor analysis.

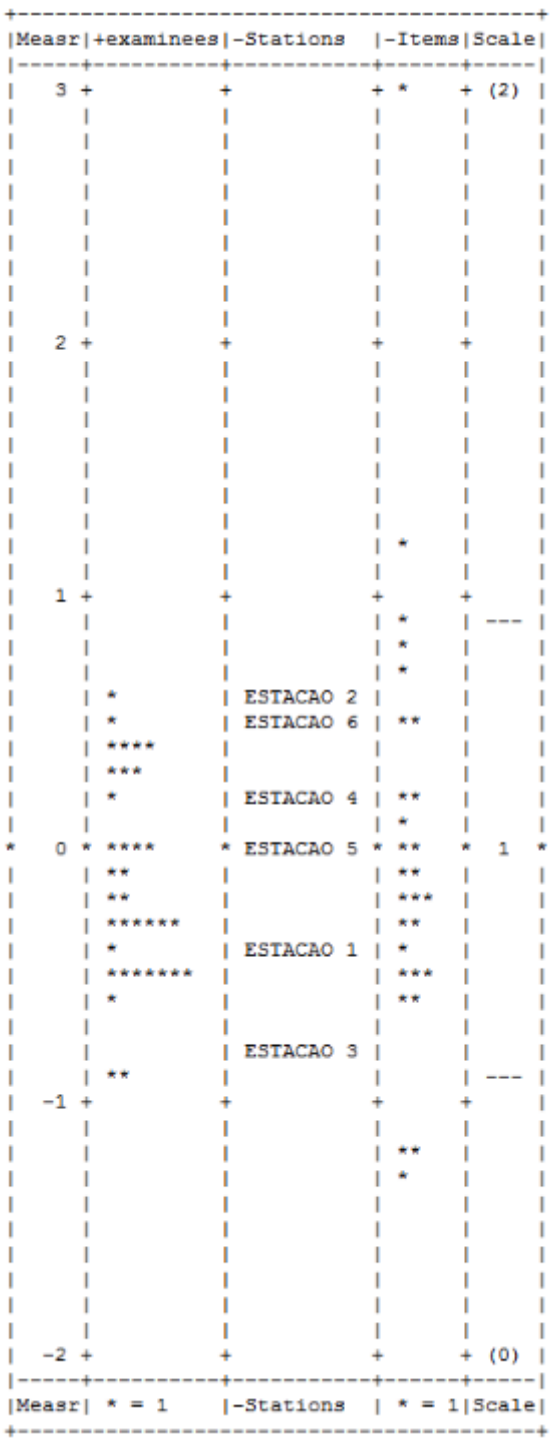

Figure 2

Graph 2. Variable map containing the distribution of the elements in each facet (examinees, stations and items) across the latent scale.

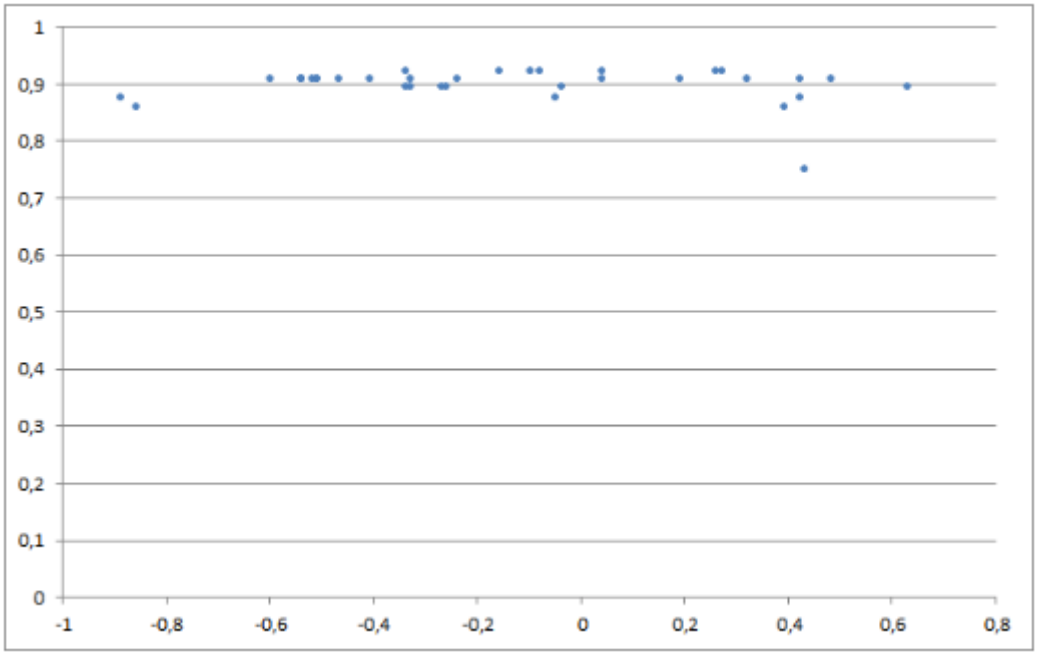

Figure 3

Graph 3. Individual precision estimates for examinees of the CCOG according to the Many-Facet Rasch Model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalfileBrazilianversionoftheCCOG.docx