Original Article

CONCEPÇÃO E VALIDAÇÃO DO PROTOCOLO DE OBSERVAÇÃO DE AULA (POA) DO PROGRAMA SEGUNDO TEMPO

CONCEPT AND VALIDATION OF THE CLASSROOM OBSERVATION PROTOCOL (COP) OF THE PROGRAMA SEGUNDO TEMPO

Antonio Jorge Gonçalves Soares[1], Maria Jullyanne Cavalcanti de Brito[2], Tony Meireles Santos[2], Amauri Aparecido Bassoli de Oliveira[3], Marcelo Silva Cardoso[4], Jose Luiz Lopes Vieira[3], Luciane Cristina Arantes da Costa[3] e Vinícius de Oliveira Damasceno[2]

[1]Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, Brasil.
[2]Universidade Federal de Pernambuco, Recife-PE, Brasil.
[3]Universidade Estadual de Maringá, Maringá-PR, Brasil.
[4]Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brasil.

RESUMO

Na oferta de uma educação esportiva em escala nacional o Programa Segundo Tempo (PST) elaborou padrões teóricos e metodológicos visando fornecer um ensino esportivo de qualidade. Para avaliar sua proposta pedagógica o PST construiu um modelo de avaliação para medir o grau de adesão dos docentes do programa ao modelo pedagógico PST. Desse modo, o objetivo do presente estudo foi elaborar o Protocolo de Observação de aula (POA), bem como estabelecer sua validade de face e de conteúdo. Para a construção do instrumento foram observadas as seguintes etapaa: a. revisão dos documentos que fundamentam o PST; b. reuniões e consultas as equipes pedagógicas; c. construção da definição operacional; d. construção dos itens; e. estudo piloto (Índice de Concordância entre os observadores ao avaliarem a mesma aula). De uma forma geral, não foram encontradas divergências entre as equipes quanto a relevância dos itens e seus ponderamentos. Além disso, a aplicação piloto apresentou Índice de Concordância de 0,71 ± 0,22. Desse modo, o POA se mostrou consistente e um excelente balizador para medir a qualidade de entrega das aulas do PST nos núcleos.
Palavras-chave: Psicometria. Validade dos testes. Avaliação educacional.

ABSTRACT

The Programa Segundo Tempo (PST) designed theorethical and methodological standards in order to provide quality in sports education on national scale. The PST built an evaluation model to measure the compliance level of teachers to the PST pedagogical model. Thus, the objective of this study was to design the Class Observation Protocol (COP), as well as to establish its face and content validity. The following steps were followed in the instrument's construction: a. review of the documents underlying the PST; B. meetings and consultations with the pedagogical teams; C. construction of the operational definition; d. construction of items; and. pilot study (Concordance Index between observers assessing the same class). In general, no differences were found between teams as to the relevance of the items and their weights. In addition, a pilot application presented a Concordance Index of 0.71 ± 0.22. Thus, the COP proved to be consistent and an excellent indicator to measure the PST's teaching quality.
Keywords: Psychometrics. Validity of tests. Educational measurement.

Introduction

The "*Programa Segundo Tempo*" (PST) (Second Half Program), as any educational program, aims at sharing a set of knowledge, values as well as sports practices and corporal activities, primarily with the population living in situations of social vulnerability, with the purpose of contributing to citizenship and the improvement in their quality of life[1]. Thus, the PST fulfills the constitutional precept that leisure and sports practices are the rights of every citizen and it is the State's duty to foster them to the entire population[2].

When providing quality sports education to its beneficiaries (children and young people), the PST carried out educational activities supported by theoretical, pedagogical and methodological debates related to sport teaching and to body activities. Thus, in the scope of the program, sports is conceived as a social good to be democratized, as a place of education

for citizenship, as a space for leisure and social inclusion and as an experience of learning motor skills and strategic game thinking (tactical intelligence)[3-5]. Therefore, educational sport is the perspective adopted by the program insofar as it is based on the principles of participation, cooperation, coeducation, integration and responsibility[6]. Although these principles may also be claimed by high-performance sports, there are differences regarding the focus and values permeating these expressions of the sports phenomenon in our society. Educational sport aims, above all, at providing opportunities for young people to reinforce democratic values, making them incorporate healthy habits in their leisure time and acquire taste for physical activity. On the other hand, high-performance sports, or institutionalized sports, typically aims at the athletes' maximum performance, the spectacle, the entertainment market, consumption and victory[7].

In this sense, having the socialization of educational sports as its goal, the PST dared to gather groups of different theoretical nuances, from 40 Brazilian universities, to formulate its pedagogical proposal of intervention. Having in mind the program`s permanent concern with the quality of the service offered to its beneficiaries, the PST's Pedagogical Team (PT) has built throughout its existence a set of didactic materials as well as training programs for teachers and monitors who work in the program's centers. Their purpose was to build conceptual, operational and methodological standards oriented to the excellence in the activities carried out in the classes[8-11].

Based on the PST's pedagogical model, we can conclude that a class that is aligned to the pedagogical actions of the Program should contemplate basic aspects such as, a) adequate planning of pedagogical actions[3,12]; b) approach of contents in a clear and safe way, contemplating the various dimensions of teaching (conceptual, procedural and attitudinal)[5,12]; c) methodological procedures based on both technical and tactical skills[3]; d) evaluation strategies based on individual and collective feedback[5]; and e) actions that promote the inclusion, adherence and satisfaction of beneficiaries in all proposed activities[4].

Therefore, besides building minimum theoretical and methodological standards to provide quality sports education, the PST knew that, if they were to operate on national scale, they should also implement an evaluation model to measure the quality of teaching offered by the Program's centers throughout Brazil.

Having this concern in mind, the Class Observation Protocol (COP) was designed as a tool to assess how much of the partner teachers' classes are in accordance to the PST pedagogical model. This information may help the program diagnose the impact of its training and its pedagogical materials (books, manuals, etc.) in the classes offered at the Program's centers in Brazil. This instrument may generate diagnoses allowing the understanding of the difficulties for didactic transposition of PST principles to the classes offered to the Program subjects. With the data about the quality of classes, we can propose more effective actions so that we can ensure that teachers are able to offer opportunities for sports experiences with the quality desired by the PST pedagogical model. This instrument could also be used as a guide for the Program teachers, and possibly for those outside it, to plan and conduct their classes. This could probably have an impact on improving the quality of interventions.

The technical qualification of the professionals using the instrument is necessary in order to measure certain variables so that they are reliable as well as the consistency level of the results obtained through the instrument used[3]. Therefore, we must verify if the COP content actually contemplates the construct that it is proposed to measure and if its items are clearly constructed under the perception of the individuals who will use it[13-16]. Thus, the objective of this study was to present an instrument to evaluate the level of teachers' compliance to the PST class proposal, as well as to establish their face and content validity.

Methods

Five members of the PET (Pedagogical Evaluation Team) and nine members of the Pedagogical Team (PT) participated in the design and management of the instrument validity stages.

The PT is the team responsible for all the PST' pedagogical management. The team is composed by researchers from the School Physical Education and Sports Pedagogy. They are responsible for: a. managing and fostering the production of the PST' theoretical and didactic materials; b. formatting and following the PST's teachers training process through Distance Learning (DL); c. promoting face-to-face training and d. monitoring the pedagogical dimension of the educational sport offer to the program's beneficiaries. On the other hand, the PET is hierarchically linked to the PST's PT, whose objective is to build a pedagogical evaluation model of the program.

Methodological Strategy

The following steps[15] were observed for the instrument construction: a. review of documents underlying the PST; b. meetings and consultations with the pedagogical teams; c. construction of the operational definition; d. construction of items; e. pilot study. Steps a, b, c, and d comprise the design, face and content validity detailed in Figure 1.
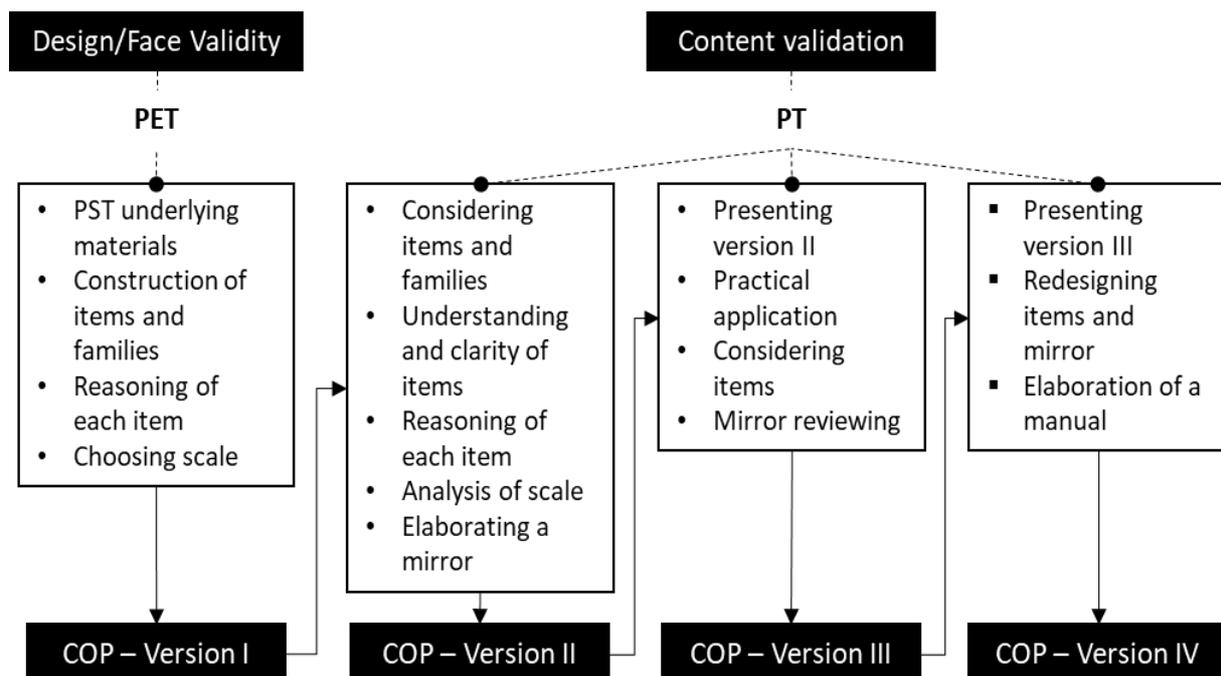

Figure 1. Detailing the instrument's validation phases
Source: The authors

Procedures

Face and Content Validity

The instrument was designed and structured in a consensus meeting between the PET and PT members. The starting point was the documents (PST's books and manuals) and a preliminary version of qualitative indicators of the lecture given by the teachers, found in Avil (3rd Generation)[17,18]. Avil means *Assessment in Loco*, a form that is used in the administrative and pedagogical monitoring of the program by the Collaborating Teams (CT). These teams

work in different regions of the country making visits to assist teachers and monitoring the functioning of the PST's centers. The 3rd Generation version includes a checking form sheet of PST classes with indicators to map the teacher's teaching actions. These indicators, besides their general character of sport and corporal activities good teaching, were anchored in the pedagogical precepts enunciated by the PST's pedagogical proposal. From this document, items were determined and others were added to make up the COP. They were grouped in families and weighed consensually in an arbitrary manner by the PET, to make up the whole (100%), culminating in the COP first version. This preliminary version was submitted to the PT members, who evaluated their items as to their relevance, clarity and comprehension, in addition to establishing weights for each one, which would later be compared with the weights suggested by the PET. They also analyzed the scale and suggested to construct a mirror detailing each item and describing the maximum score, resulting in the COP Version II.

This new version was presented to the PT members and implemented by them through the evaluation of two video classes. Then, the PT analyzed item by item and had to classify them as adequate or inadequate in relation to how the item was described, as well as the mirror detailing and the maximum score. When classifying any of these aspects as inadequate, they should make a remark justifying it; if all PTs members agreed with the judgments, the modification was accepted. All data were collected on a digital platform (*Surveymonkey.com*).

The COP Version III was presented to the PT, which re-adjusted items and the mirror. A manual was also prepared, aiming at enhancing the clarity of what should be evaluated in each item, resulting in the COP IV Version.

Pilot Study

A distance learning course (DL) was elaborated as the pilot study in order to train PT and CT members in the COP use. Twenty-nine PT and CT teachers watched a model class and, at the same time, completed the COP. Subsequently, they were asked to read the COP instruction manual, and then to watch two 45-minute lessons each (recorded at the PST center - Recife - Pernambuco - Brazil) with simultaneous completion of the COP. The DL course followed the chronology shown in Figure 2. The project was approved by the UFPE (Federal University of Pernambuco) Ethics and Research Committee under number 040336.
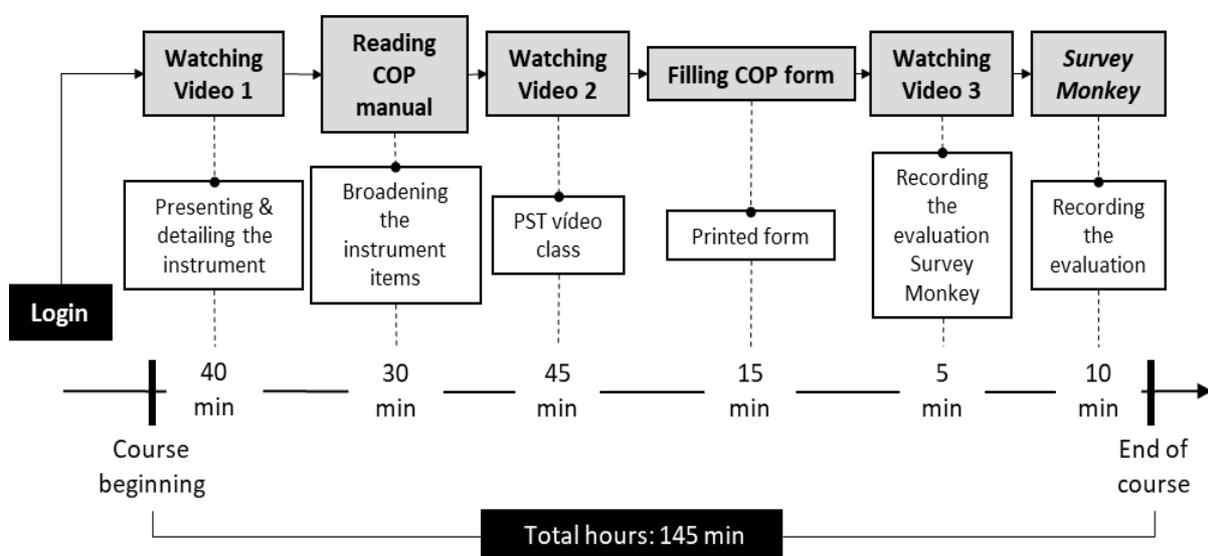

Figure 2. Phases of DL capacitation
Source: The authors

The classes used in the course were previously observed by PET members who used the COP in order to metrify and establish a reference for the classes. A camera (*HandyCam* HR-11SR SONY, FULL HD, 120 GB *Hard disk*, USA) and a tripod (Sony® VCT-60AV) was used to record the classes. The camera was placed in a favorable position so that it could visualize everything in the space used in class. A clip-in microphone was used (Saramonic® Professional Microphone Lapela Wireless Sr-wm4c) to capture the teacher's speech. The videos were edited in *Adobe Premier Pro CC 2015*.

Statistical analysis

Descriptive statistics (mean, standard deviation, coefficient of variation and 95% CI) were used for statistical analysis. The Chi-Square test was used to verify the difference of the scores of each inter-rater indicator. The concordance index was also used to verify agreement among the same class observers. The classification proposed by Camargo and Sentelhas (1997)[19] was used to classify the concordance index.

The *SPSS® 22 for Windows® (IBM), GraphPad Prisma® 5 for Windows®* software was used for both phases. The significance level adopted was $p < 0.05$.

Results

The COP underwent changes in both its structure and its content during the various its construction and validation stages. Table 1 summarizes these changes

Table 1. Detailing the main adjustments made in the COP versions

Structure	Versions			
	Version I - COP	Version II - COP	Version III -COP	Version IV – COP
IteMs	30	28	28	22
Wording	Long/affirmative and negative	Long/affirmative and negative	Long/affirmative	Short/affirmative
Scales	0 to 2	0 to 2 and "Not applicable"	0 to 2 and "Not applicable"	0 to 2 and "Not applicable"
Mirror	-	Item detailing and meaning of scale 2	Item detailing and meaning of scale 2	Item detailing and meaning of scale 0, 1 and 2 and "not applicable"
Manual	-	-	-	Manual elaboration

Source: The authors

The scale adopted for the COP ranges from 0 to 2 points based on a consensus decision among the PET members. The binomial approach (1 or 2) with an option for the item non-presence (0) was made to simplify the observation strategy given its later application in the evaluation process, thus minimizing the phenomenon of 'paradox of choice '. The grade 0 shows that the teacher did not achieve/execute the observed aspect, grade 1 shows that he/she achieved/executed it partially and grade 2 shows that he/she achieved/executed it properly. Furthermore, it is also present in some of the specific indicators a fourth option named "**Not applicable**", which should be checked when the specific configuration of the class (the type of class taught) does not require that aspect to be observed.

Table 2 shows the results referring to COP Version II when submitted to the PT, in which the item, its description in the mirror and the detail of the maximum score were evaluated as adequate or inadequate.

Table 2. Detailing the evaluation carried out by the PT – COP Version II

Indicator Families	Item		D.I		D.D		Adjustments
	Ad.	Inad.	Ad.	Inad.	Ad.	Inad.	
Planning Indicators							
1. Class Objectives	1.0	0.0	1.0	0.0	1.0	0.0	-
2. Coherent sequencing of activities	1.0	0.0	0.9	0.1	1.0	0.0	-
3. Occupation of the available physical space	1.0	0.0	1.0	0.0	1.0	0.0	-
4. Use of materials x number of students	0.9	0.1	1.0	0.0	0.8	0.2	Inclusion of "not applicable"
5. Compatibility of activities per age group	0.5	0.4	0.5	0.5	0.6	0.4	Words: age group x motor level
6. Use of materials x motor experiences	0.5	0.5	0.6	0.3	0.6	0.3	Item excluded
Content Indicators							
7. Confidence and clarity in teaching the contents	0.7	0.3	0.7	0.3	0.8	0.2	Exclusion of the word "clarity"
8. Procedural information	0.6	0.4	0.5	0.5	0.8	0.2	Joining with item 9
9. Conceptual Information	1.0	0.0	0.9	0.1	0.9	0.1	Joining with item 8
10. Ethical and moral values	0.9	0.1	1.0	0.0	1.0	0.0	-
Methodological procedures Indicators							
11. Respectful relationship with students	0.9	0.1	1.0	0.0	0.9	0.1	Exclusion of "respectful"
12. Class organization	1.0	0.0	0.9	0.1	1.0	0.0	-
13. Leadership in class	0.6	0.4	0.6	0.4	0.6	0.4	Defining "leadership"
14. Challenges/problems to acquiring skills	0.9	0.1	0.8	0.2	1.0	0.0	Item excluded
15. Self-organization	1.0	0.0	1.0	0.0	1.0	0.0	Joining with item 12
16. Teaching the sport from the game goal	1.0	0.0	0.9	0.1	0.9	0.1	Inclusion of "not applicable"
17. Use of mini-games	0.9	0.1	0.9	0.1	0.8	0.2	Inclusion of "not applicable"
18. Presence of lines and big games	1.0	0.0	1.0	0.0	0.9	0.1	Replace presence for absence
19. Period of experience in the activities	1.0	0.0	0.9	0.1	0.9	0.1	Joining with item 18
20. Teaching the sport from the partial method	0.9	0.1	0.7	0.3	0.9	0.1	Joining with item 16
21. Strategic and tactical behavior	0.9	0.1	1.0	0.0	1.0	0.0	Inclusion of "not applicable"
22. Redesigning or adaption of activities	1.0	0.0	0.8	0.2	1.0	0.0	Inclusion of "not applicable"
Evaluation Indicators							
23. Individual Feedback	1.0	0.0	0.9	0.1	1.0	0.0	Definition of "feedback"
24. Collective Feedback	1.0	0.0	0.8	0.2	1.0	0.0	Definition of "feedback"
25. Fostering auto enhancement in the activities	1.0	0.0	0.9	0.1	1.0	0.0	Item excluded
26. Final analysis of the class	0.9	0.1	0.9	0.1	0.9	0.1	-
Students adherence/inclusion indicators							
27. Contentment e pleasure with the class	0.8	0.2	0.7	0.3	0.9	0.1	Exclusion of "pleasure"
28. Participation of those less skillful	0.7	0.3	0.8	0.2	0.8	0.2	Change "less skillful" to "more vulnerable to exclusion"

Remarks: D.I. = Description of item in the mirror, D.D. = Description of the detailing of maximum score Ad. = Adequate, Inad. = inadequate
Source: The authors

As for the items weights of each family, the PET assumed the mean values and the respective suggestions to decrease, maintain or increase the proposed initial weight, as listed in Table 3.

Table 3. Structure of the last COP version and the means of weights to each specific indicator

Indicator Families	Weight	Mean (CI ± 95%)	PT Suggestion
Planning Indicators	20%		
1. Class Objectives		1.5 ± 2.0	Increase weight
2. Coherent sequencing of activities		1.8 ± 2.2	Decrease weight
3. Occupation of the available physical space		1.5 ± 2.3	Increase weight
4. Use of materials		1.3 ± 2.8	Decrease weight
5. Compatibility of activities per age group		1.7 ± 2.1	Increase weight
Content Indicators	20%		
6. Confidence in teaching the contents		1.7 ± 1.7	Decrease weight
7. Suitable language for teaching		1.8 ± 1.8	Decrease weight
8. Conceptual and Procedural information on contents		2.6 ± 1.5	Increase weight
9. Ethical and moral values		1.5 ± 2.0	Increase weight
Methodological procedures indicators	30%		
10. Rapport with the students		1.5 ± 2.0	Increase weight
11. Class organization		1.5 ± 2.0	Increase weight
12. Leadership in class		1.3 ± 2.6	Decrease weight
13. Acquisition of technical skills		1.7 ± 2.8	Decrease weight
14. Teaching the game		-2.3 ± 2.0	Decrease weight
15. Period of experience in the activities		1.8 ± 2.4	Decrease weight
16. Strategic and tactical behavior		2.8 ± 3.4	Decrease weight
17. Redesigning the activities		1.7 ± 2.8	Decrease weight
Evaluation Indicators	15%		
18. Individual feedback		1.7 ± 2.5	Decrease weight
19. Collective feedback		1.3 ± 1.8	Increase weight
20. Final analysis of the class		2.2 ± 2.7	Increase weight
Students adherence/inclusion indicators	15%		
21. Contentment and pleasure with class		2.2 ± 3.6	Increase weight
22. Inclusion		2.3 ± 3.2	Increase weight

Remarks: CI = Confidence Interval, PT = Pedagogical Team.
Source: The authors

The overall mean of the Concordance Index was 0.71 ± 0.22, with a confidence interval of (0.61-0.81). Thus, according to the weights suggested by the PET and the mean value proposed by the PT for each specific indicator, there was no call for changes in the percentage value for the Content family. There was a suggestion for a weight reduction for the Methodological Procedures, Evaluation and Adhesion families, although the mean did not present a significant difference for the percentage weight proposed by the evaluation team. Therefore, there was no change in the original proposal.

Each of the specific indicators had a different weight, according to the degree of importance the PT assigned to them. The set of specific indicators measures the compliance of the class to each family of indicators in a centesimal scale, as shown in Table 3.

Finally, the COP Version IV presents 22 items, which were called *specific indicators*. They were distributed among five indicator families, in which each specific indicator expresses elements that must be present in a PST class. COP Version IV is available at (https://www.dropbox.com/s/kskdi4rhax5h0g6/POA%20-%20Formul%C3%A1rio%20Final%20-%20Completo.pdf?dl=0). There is also the COP Handbook (https://www.dropbox.com/s/wvssq55w9udl74x/Manual%20PST%20-%20Protocolo%20de%20Observacao%20de%20Aula%20-%20vers%C3%A3o% 20final.pdf? Dl = 0), which shows, in a detailed and contextualized way, all the instrument structure through images, examples and theoretical background based on the PST support materials, reporting item by item, from its description to specification of when each grade of the scale should be checked.

To obtain the final indicator of a class, which translates the teachers' compliance degree to the PST methodology, the values must be transposed into a spreadsheet oriented to the analysis of this data and that was based on the weights assigned to each item. The values generated are classified as satisfactory (≥ 70%) or unsatisfactory (<70%).

The calculation of the class compliance to each family, the scores assigned to each item should be multiplied by their **weight**. These values will be added and later divided by the sum of the maximum score multiplied by the weight (Equation 1).

$$x = \frac{(Observed\ score\ x\ weight)}{\sum(Maximum\ score\ x\ weight)} \times 100$$

The compliance with the class general quality will be o calculated by Equation 2.

$$Class\ General\ Qaulity = \frac{[(PS \times 20) + (CS \times 20) + (MPS \times 30) + (EvS \times 15) + (SAS \times 15)]}{5}$$

Where: PS – Planning score; CS – Contents score; MPS – Methodological Planning score; EvS – Evaluation score; SAS – Students Adherence/Inclusion Score

Table 4 lists the Concordance Index values among the evaluators observing the same video lesson in the pilot study.

Table 4. Concordance Index among the evaluators for each specific indicator

Indicator Families	Median	CI	Class
Planning Indicators			
1. Class objectives	2 (2 - 2)	0.93	Great
2. Coherent sequencing of activities	2 (2 - 2)	0.9	Great
3. Occupation of the available physical space	2 (1 -2)	0.48	Bad
4. Use of materials	2 (1,75 - 2)	0.76	Very good
5. Compatibility of activities per age group	2 (2 - 2)	0.93	Great
Content Indicators			
6. Confidence in teaching the contents	2 (2 - 2)	0.93	Great
7. Suitable language for teaching	2 (1 - 2)	0.9	Great
8. Conceptual and Procedural information on contents	2 (1 - 2)	0.72	Good
9. Ethical and moral values	2 (2 - 2)	0.52	Bad
Methodological procedures indicators			
10. Rapport with the students	2 (2 - 2)	0.93	Great
11. Class organization	2 (2 - 2)	0.93	Great
12. Leadership in class	2 (1 - 2)	0.24	Terrible
13. Acquisition of technical skills	2 (1 - 2)	0.63	Moderate
14. Teaching the game	2 (2 - 2)	0.83	Very good
15. Period of experience in the activities	2 (2 - 2)	0.86	Very good
16. Strategic and tactical behavior	2 (1 - 2)	0.62	Average
17. Redesigning the activities	2 (#)	0.31	Terrible
Evaluation Indicators			
18. Individual feedback	2 (1 - 2)	0.58	Bad
19. Collective feedback	2 (2 - 2)	0.93	Great
20. Final analysis of the class	2 (1 - 2)	0.76	Very good
Students adherence/inclusion indicators			
21. Contentment and pleasure with class	2 (1 - 2)	0.28	Terrible
22. Inclusion	2 (1 - 2)	0.72	Good

CI: Confidence Interval
Source: The authors

In general, it can be seen that since this first stage of construction, the COP presents a theoretical and conceptual coherence with the PST's curricular parameters. PT and PET members found no differences regarding either the items relevance or their weights.

Discussion

This study intended to present the COP in conjunction with the first procedures aimed at establishing its psychometric qualities. As this instrument is a qualifying test, it is essential that the content and data derived from it are reliable and measure in a consistent way what it intends to measure[13,15]. Precautions must be taken when designing any instrument[15], thus, the necessary phases for the construction and validity of COP content were carefully considered and based on the existing literature[15,20].

It is well-known in the scientific world that it is preferable to validate an existing instrument than to create a new one[15]. Therefore, before designing the COP, we were concerned to look for observational instruments that had been created with the purpose of evaluating the quality of teaching[21-23]. Among such instruments, the one that came closest to the PST classroom reality was the *Assessing Quality Teaching Rubrics* (AQTR), which was designed to evaluate the teaching practices of pre-service physical education teachers[23]. However, its items address more general questions about how a teacher should conduct a lesson, which would also be valid for a class in the PST model, but not enough, given that the AQTR does not clearly address aspects such as: dealing with ethical and moral values, a respectful relationship between teacher and student (absence of discriminatory language), the teaching of a game from functional structures and inclusion of those students most vulnerable to exclusion, and students' satisfaction with the class.

The emphasis in these aspects would be the great differential between a PST class in relation to other classroom settings aimed at the common teaching of sports techniques and corporal practices[9,10]. Therefore, we chose to build a new instrument that was specifically developed for the Program, encompassing all the aspects we considered essential to be evaluated if the class is aligned with the PST standards.

After its construction, we applied the instrument to a pilot sample, as recommended[15], in order to verify the agreement between the observers for each item during the observation of a given class. More than half of the items showed a rating of "Good" and "Very Good" with the exception of some items ranked "Moderate."

Having the disagreements found in some items in mind, we knew that one of the aspects that can influence the result of a test is the competence of the evaluators who will use the instrument[13]. Therefore, the technical qualification of these personnel is essential to ensure consistent results[15]. Thus, prior to starting the evaluation of the class itself, the evaluators were submitted to a training that would lead them to know the COP in detail, from its objective to the final result that would be generated. This way, besides aggregating the phases of data collection, the DL course was created to train its participants through an explanatory video lesson followed by reading the COP Manual.

When we talk about observation, we must also consider factors that can influence the response given by the observers and that end up generating different levels of error. Among them, we can mention benevolence, harshness, central tendency errors as well as the halo effect. In general, some of these errors occur when the observer already knows or has heard of the individual who will be observed, when he is influenced by racial and/or philosophical prejudices, and thus, is afraid of being too severe or too benevolent in his evaluation[13,15]. The way we found to try to minimize these errors was by alerting observers during the training, about their existence, as a way to make them critically analyze their decision-making.

Considering also that another source of measurement error may be directly related to the imprecision of the instrument[15], the items that presented disagreement were analyzed and described in more detail in the manual, in order to minimize the divergence between the observers evaluation during its use.

In general, considering that this study aimed to introduce the COP to the scientific community, it is important to emphasize that the next steps are still necessary so that the instrument can finally be used with the necessary scientific rigor[13,15,20]. The next step concerns the COP application to a target sample. From this last application, we will be able to verify if the instrument is in fact valid and reliable to be used by the Program and by those interested in evaluating classes focused on the teaching of the educational sport and the body practices. Therefore, we have an ongoing study that aims at establishing the validity of concurrent criteria (comparing observers' results with the gauge) and intra- and inter-observer reliability (analyzing the data repeatability).

Conclusions

The COP has shown to be consistent to measure not only the teachers' degree of compliance to the pedagogical practices set forth by the PST, but also if the instrument assumes the heuristic characteristic of being a guide for the teachers themselves to think about their teaching actions. This instrument will be able to fulfill this dual purpose and add value to the quality of the program: to obtain data from on-site evaluations as well as the implications for future training, and to contribute to the development of a common language among program teachers, who will glimpse the essential actions that must be present in their classes in a practical way.

References

1. Hansen FR, Perim GL, Oliveira AAB. Apresentação. In: Oliveira AAB, Perim GL, organizadores. Fundamentos Pedagógicos do Programa Segundo Tempo: da reflexão à prática. Maringá: Eduem; 2009, p. 1-16.

2. Brasil. Constituição da República Federativa do Brasil de 1988. [acesso em: 30 set 2017]. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm.

3. Greco PJ, Silva SA, Santos LR. Organização e Desenvolvimento Pedagógico do Programa Segundo Tempo. In: Oliveira AAB, Perim GL, organizadores. Fundamentos Pedagógicos do Programa Segundo Tempo: da reflexão à prática. Maringá: Eduem; 2009, p. 163-206.

4. Melo VA, Brêtas A, Monteiro MB. Fundamentos do lazer e da animação cultural. In: Oliveira AAB, Perim GL, organizadores. Fundamentos Pedagógicos do Programa Segundo Tempo: da reflexão à prática. Maringá: Eduem; 2009, p. 45-72.

5. Palma, MS, Valentini NC, Petersen R, Ugrinowitsch H. Estilo de Ensino e Aprendizagem Motora: Implicações para a prática. In: Oliveira AAB, Perim GL, organizadores. Fundamentos Pedagógicos do Programa Segundo Tempo: da reflexão à prática. Maringá: Eduem; 2009, p. 89-114.

6. Tubino M. O que é esporte. Tatuapé: Brasiliense; 1993.

7. Brasil. Lei nº 9.615 de 24 de Março de 1998. [acesso em: 30 set 2017]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L9615consol.htm

8. González FJ. Ensino dos esportes. In: González FJ, Darido SC, Oliveira AAB. Práticas Corporais e Organização do Conhecimento. Maringá: Eduem; 2014, p. 29-60.

9. Oliveira AAB, Perim GL. Fundamentos Pedagógicos do Programa Segundo Tempo: da reflexão à prática. Maringá: Eduem; 2009.

10. Oliveira AAB, Perim GL, Fundamentos Pedagógicos para o Programa Segundo Tempo. Porto Alegre: Brasília: Ministério dos Esportes; 2008.

11. Greco PJ, Conti G, Morales JCP. Manual de Práticas para a Iniciação Esportiva no Programa Segundo Tempo. Maringá: Eduem; 2013.

12. Melo JP, Dias JCNSN. Fundamentos do Programa Segundo Tempo: entrelaçamentos do esporte, do desenvolvimento humano, da cultura e da educação. In: Oliveira AAB, Perim GL, organizadores. Fundamentos Pedagógicos do Programa Segundo Tempo: da reflexão à prática. Maringá: Eduem; 2009, p. 17-44.

13. Thomas JR, Nelson JK, Silverman SJ. Medidas de Variáveis de Pesquisa. In: Thomas JR, Nelson JK, Silverman SJ. Métodos de Pesquisa em Atividade Física. 6.ed. São Paulo: Artmed Editora; 2011, p. 213 - 33.

14. Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. Am J Med, 2006;119(2):166.e7-.e16.

15. Hutz CS, Bandeira DR, Trentini CM. Psicometria - Coleção de Avaliação Psicológica. São Paulo: Artmed Editora; 2015, p.177.

16. DeVon, HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. J Nurs Scholarsh 2007;39(2):155-164.

17. Streiner DL, Norman GR, Cairney J. Health measurement scales: a pratical guide to their development and use. 5. ed. United Kingdom: Oxford, 2015.

18. Kravchychyn C, Oliveira AAB. Esporte Educacional no Programa Segundo Tempo: uma construção coletiva. J. Phys. Educ 2016; 27(1):1-18.

19. Camargo AC, Sentelhas PC. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo, Brasil. Revista Brasileira de Agrometeorologia 1997;5(1):89-97.

20. Pasquali L. Psychometrics. Rev Esc Enferm USP 2009;43(Esp):992-999.

21. Hill HC, Blunk ML, Charalambous CY, Lewis JM, Phelps GC, Sleep L, et al. Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. Cogn Instr 2008;26(4):430-511.

22. Burry-Stock JA, Oxford RL. Expert science teaching educational evaluation model (ESTEEM): Measuring excellence in science teaching for professional development. Journal of Personnel Evaluation in Education 1994;8(3):267-97.

23. Chen W, Hendricks K, Archibald K. Assessing pre-service teachers' quality teaching practices. Educ Res Eval 2011;17(1):13-32.

Author address: Antonio Jorge Gonçalves Soares. Rua Alzira Côrtes, 5. Botafogo, Rio de janeiro, C.E.P. 22260-050. E-mail: ajgsoares@gmail.com