

Angélica Pilar Silva Ríos^{1,2} 

Manuel Nibaldo del Campo Rivas³ 

Patricia Katherine Kuncar Uarac¹ 

Víctor Antonio Calvo Sprovera¹ 

Reliability of a script agreement test for undergraduate speech-language therapy students

Validación de constructo y reproducibilidad de una prueba de concordancia de guiones para estudiantes de pregrado en fonoaudiología

Keywords

Reliability
Reproducibility of Results
Higher Education
Speech-Language Pathology
Professional Competence

Descriptores

Confiabilidad
Reproducibilidad de los Resultados
Educación Superior
Fonoaudiología
Competencia Profesional

ABSTRACT

Purpose: To estimate the reliability of scripts designed for undergraduate Speech-Language Therapy students. **Methods:** A descriptive cross-sectional study was carried out. Qualitative variables were summarized by frequency or proportion and quantitative through means (CI 95%). Reliability was estimated through Cronbach's α coefficient, and inter-rater agreement was determined using Fleiss's Kappa index. The analytical tests considered a significance level of $p < 0.05$. **Results:** 80 scripts organized in four areas of speech-language therapy were validated by 41 speech-language pathologists. The average experience of the professionals was 17.1 years. The reliability of the corpus was $\alpha: 0.67$ (min= 0.34; max: 0.84), and the inter-rater agreement $\kappa: 0.29$ (min: 0.07; max: 0.45). **Conclusion:** The corpus's reliability scores were similar to those reported by previous studies in different health professions. Having validated strategies aimed at developing proficiency and supporting classic training actions in undergraduate courses will contribute to increasing the quality of future health professionals.

RESUMEN

Objetivo: Estimar la confiabilidad y reproducibilidad de un corpus de scripts diseñado para el pregrado de fonoaudiología. **Método:** Estudio observacional de tipo descriptivo y temporalidad transversal. Se estimó la validez de constructo a partir del coeficiente α de Cronbach y la reproducibilidad con el índice Kappa de Fleiss. Las pruebas analíticas consideraron un nivel de significancia $p < 0.05$. **Resultados:** Se creó un corpus de 80 scripts organizados en 4 áreas de la fonoaudiología el que fue validado por 41 fonoaudiólogos. La experiencia promedio de los profesionales fue de 17.1 años (Std. Err: 2.4; IC 95%: 11.7-22.6). La confiabilidad del corpus fue $\alpha: 0.67$ y el acuerdo interevaluador, $\kappa: 0.29$. **Conclusión:** Los puntajes de confiabilidad y reproducibilidad del corpus creado fueron similares a los reportados por estudios previos en otras profesiones de la salud. Contar con estrategias validadas que se orienten al desarrollo de competencias y complementen las acciones formativas, contribuirá a incrementar la calidad en la formación de futuros profesionales de la salud.

Correspondence address:

Angélica Pilar Silva Ríos
Escuela de Fonoaudiología, Facultad de Salud, Universidad Santo Tomás – UST
Av. Uno norte #3041, 5° piso, Carrera de Fonoaudiología, Viña del Mar, Valparaíso, Chile, Código postal: 2561694
E-mail: asilvar@santotomas.cl

Study conducted at Universidad Santo Tomás – Chile.

¹ Escuela de Fonoaudiología, Facultad de Salud, Universidad Santo Tomás – Chile.

² Centro Interdisciplinario de Innovación Educativa, Universidad Santo Tomás – Chile.

³ Escuela de Fonoaudiología, Facultad de Ciencias de la Salud, Universidad Católica Silva Henríquez - Santiago, Región Metropolitana, Chile.

Financial support: Proyecto de Innovación Educativa, Centro Interdisciplinario de Innovación Educativa, Universidad Santo Tomás, Chile (11400007).

Conflict of interests: nothing to declare.

Received: April 28, 2022
Accepted: June 09, 2023



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The training of health professionals has become more complex to the extent that disciplinary and teaching knowledge has expanded, involving teaching-learning processes supported by evidence, centered on the student, and accompanied by the use of technologies, innovation, and high scientific standards⁽¹⁾. Students must develop proficiency involving knowledge, skills, behaviors, and values that allow them to approach users in diverse and uncertain contexts⁽²⁾.

Among the skills that students must develop in their academic training, there is clinical reasoning (CR), a relevant component for developing competencies that allow inferring a diagnosis and choosing the most appropriate routes of action for each user⁽³⁾. In the case of Speech-Language Therapy, a discipline in charge of addressing and preventing communication and swallowing disorders⁽⁴⁾, undergraduate training requires the student to develop complex skills⁽⁵⁾, among which is the CR, a challenging competence in terms of professional training⁽⁶⁾. Despite the existence of heterogeneous definitions, it is accepted that CR implies using skills such as: analysis, problem-solving, contextual reasoning, and pattern recognition⁽⁷⁾. Such tasks seem to happen automatically in the case of experts. However, novice students require training to achieve them⁽⁸⁾.

This way, CR is considered a fundamental skill in the approach to users, which must be incorporated into the study plans of health professionals. Within the strategies described in the literature, using scripts (clinical scripts) has offered a framework to understand how thinking is organized when training CR⁽⁹⁾.

The scripts respond to schemes associated with sequences of events that occur in a particular order⁽¹⁰⁾, organizing themselves as a set of attributes that are more or less likely to happen and that contain symptoms and signs that are linked, forming generic structures that can represent any disease⁽⁹⁾. This is how the scripts represent a decision-making model based on hypothesis generation. These are raised based on user signs and data that allow confirming or rejecting these assumptions through psychological or mental processes in which theoretical knowledge and accumulated practical experience participate⁽⁹⁾.

In this line, different studies have shown the effectiveness of scripts as a tool for teaching CR skills⁽¹¹⁾ and as a valid and reliable evaluation tool⁽¹²⁾.

It has been observed that the evaluation methods traditionally used to measure competencies are limited in assessing CR⁽¹³⁾. This is how the script agreement test has been described as an appropriate tool for training the CR, since it allows exercising decision-making and data interpretation in situations of uncertainty⁽¹⁴⁾. The scripts enable students to assess the change in the probability of a hypothesis, contrasted with new information, in scenarios focused on making diagnostic, study, or intervention decisions. These hypotheses are evaluated on a 5-point Likert scale, with positive, neutral, or negative response options⁽¹⁵⁾. A methodology challenge points out that there is still little information on cut scores or their correlation

with knowledge scores at the expert level. On the other hand, the literature highlights that experienced professionals tend to avoid extreme responses offered by the script, increasing test scores⁽¹⁶⁾, with the consequent probability of generating biases in its construct⁽¹⁷⁾. The script agreement test has been used in different health professions as a strategy for CR training; however, no studies validate its use in Speech-Language Therapy, which requires the development of scripts relevant to the profession and the need to determine specific psychometric properties. This research aimed to estimate the reliability of a corpus of clinical scripts designed for undergraduate Speech-Language Therapy students.

METHODS

The Scientific Ethics Committee of Santo Tomás University approved the research with code 153.20. Each participant signed an informed consent to enter the project.

A descriptive study of transverse cut. The universe was made up of professionals who worked in various clinical fields of speech-language therapy and had a degree in Speech-Language Therapy granted by a Chilean or foreign university. Sampling was done for convenience following the snowball strategy. The inclusion criteria were: having a minimum of 3 years of professional experience in the disciplinary area, having a Master's or Doctor's degree, and being authorized as an individual provider in the National Registry of the Health Superintendence (HS). The exclusion criteria were: working preferably in management or teaching tasks (without clinical activity) and being currently inactive in labor terms due to retirement. The necessary number of experts was selected according to the recommendation of the literature for the validation of clinical scripts, estimating the incorporation of at least ten professionals per area⁽¹⁸⁾. The intervened career contemplates four core areas, namely: Child-Youth Communication and Language; Voice and Orofacial Motor Skills; Cognition and Communication in Adults and the Elderly; Audiology and Vestibular, for which the minimum n was projected in 40 speech-language therapists who were contacted at the national level through the sampling technique mentioned earlier, corroborating an intermediate professional experience, and with a gender distribution consistent with the professional profile of the practitioners in the country.

In procedural terms and for creating the corpus, eight professionals (two per area) were invited to write the scripts to train and evaluate CR in diagnostic, study, and intervention tasks. These experts received synchronous training of three chronological hours, made online, through the Microsoft Teams platform.

The training was facilitated by an academic with a doctorate in education and applied experience in the methodology, specifically for undergraduate and postgraduate medical training. Two speech-language therapists with clinical and educational experience advised the instance, belonging to the original undergraduate program. A theoretical presentation was made, followed by a workshop for creating scripts in virtual rooms according to area or field of training. Subsequently,

the products were socialized, and feedback was provided by the facilitator to solve any procedural or theoretical doubts concerning the preparation of scripts. Each professional participating in the training was responsible for creating ten scripts.

Following the methodology reported in the literature, the population that would use the clinical scripts was determined⁽¹⁹⁾; in this case: students who completed the seventh and eighth semesters of the Speech-Language Therapy course at Santo Tomás University, at the Viña del Mar, Santiago, and Puerto Montt campuses, during the year 2021. For this research, the students did not use the scripts; however, their characterization is important, given their relevance as recipients of the final product. The identification of key contents for the corpus was made by analyzing learning results (LR) of subject programs⁽²⁰⁾ that contributed to the four areas previously declared.

A corpus of 80 scripts was created, organized into four training areas, meaning 20 per area. The scripts for each area were classified as follows: eight diagnostic scripts, two study scripts, and ten intervention scripts.

Once the scripts were generated in the established format, they were submitted to be reviewed by the research team made up of two speech-language therapists with master's degrees and experience in teaching and educational innovation, together with the expert with a doctorate degree from the medical training area, to standardize the texts and adapt them to the purposes of the study.

For the response and validation of the clinical scripts (80 in total), each Speech-Language therapist received a booklet with instructions and had to respond within a maximum of 2 weeks, depending on their area of expertise.

The responses of the speech pathologists were tabulated in Microsoft Excel. The response frequency was determined for descriptive purposes by the script to calculate the score later, transformed score, and final score according to the previously reported methodology. Quantitative variables were summarized by measures of central tendency and dispersion, considering a CI of 95%. According to the recommendation of the literature⁽²¹⁾, the internal consistency was estimated from Cronbach's α coefficient, and inter-observer reliability was calculated with Fleiss's Kappa coefficient, considering a significance level of $p < 0.05$ ⁽²²⁾. Responses were stratified by area, type of script, and type of response to control confounding variables. The incomplete scripts that presented very low response frequencies according to the outliers method were eliminated from the corpus⁽²³⁾.

The statistical analysis was performed in Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC. software.

RESULTS

Forty-one speech pathologists answered the booklets delivered by area, with a minimum age of 28 years and a maximum of 61. Regarding the distribution by gender, 78% of the women were observed in the field of Child-Youth Communication and Language, and 60% of the males in Voice and Orofacial Motor Skills. 98% of the informants had a master's degree and 2% a doctorate. The average professional experience was recorded in a range of 10.6 to 17.2 years (Table 1).

The final scores added and stratified by area and type of script reveal consistent response trends among practitioners (Figure 1). It is evident that the frequency of the added scores is usually adjusted to the type of response expected by the script: positive, neutral, or negative. In the case of the areas of Voice and Orofacial Motor Skills, together with Cognition and Communication in Adults and the Elderly, a shift towards positive values of neutral responses was noted (Figure 2E, H) for study-type scripts, and in the case of diagnosis (Figure 2J), and intervention and Audiology and Vestibular scripts (Figure 2L). In the case of the Voice and Orofacial Motor Skills area, in the same way, there is a movement of the negative responses to the right for the diagnostic case (Figure 2D) and study (Figure 2E) and the same situation was observed in the study scripts in the Audiology and Vestibular area (Figure 2K).

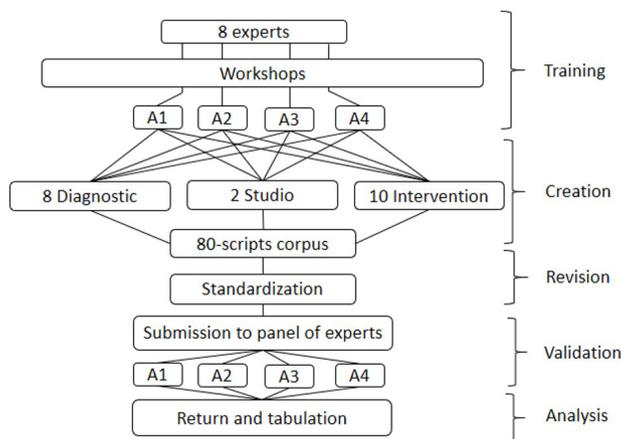
The internal consistency of the corpus was $\alpha = 0.67$. When analyzing by area and type of Script, low reliability was observed ($\alpha = 0.60-0.70$) for the diagnostic items in the four areas and the intervention scripts in the case of Cognition and Communication in Adults and the Elderly. It should be noted that insufficient reliability was recorded in the Audiology and Vestibular area case for the study scripts ($\alpha = 0.34$). On the other hand, the scripts of the same type in the remaining areas showed acceptable reliability ($\alpha = 0.80-0.90$) (Table 2).

The agreement between evaluators was estimated at $\kappa = 0.29$ (acceptable) for the corpus. When stratifying according to the variables of area and type of script, a slight concordance was observed in the areas of Voice and Orofacial Motor Skills ($\kappa = 0.07$), Cognition and Communication in Adults and the Elderly ($\kappa = 0.08$). The rest of the items presented an acceptable agreement ($\kappa > 0.20$ and < 0.41) and moderate agreement for the diagnostic scripts in Audiology and Vestibular ($\kappa = 0.45$) (Table 3).

Table 1. Sociodemographic characteristics of the experts by area

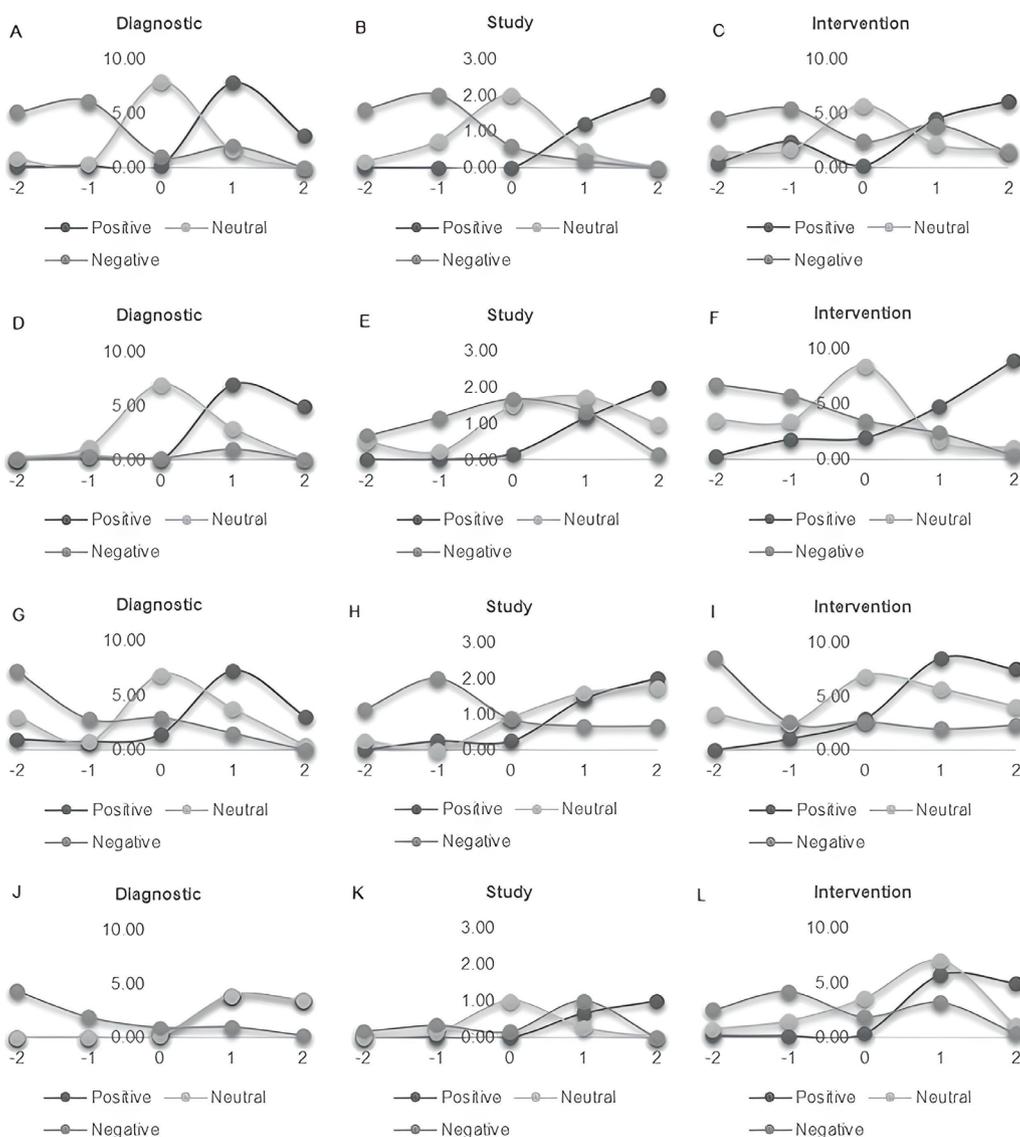
Area	Experts		Age		Experience			Gender	
	n	Average	Standard Error	Reliability Interval 95%	Average	Standard Error	Reliability Interval 95%	Male	Female
Child-Youth Communication and Language	11	40.7	2.56	35 - 46.45	17.18	2.45	11.7 - 22.66	22%	78%
Voice and Orofacial Motor Skills	10	35.4	2.16	30.51 - 40.29	11.2	2.22	6.17 - 16.22	60%	40%
Cognition and Communication in Adults and the Elderly	10	34	0.82	32.15 - 35.8	10.6	0.76	8.87 - 12.32	40%	60%
Audiology and Vestibular	10	34.8	1.38	31.66 - 37.94	10.6	1.31	7.62 - 13.58	40%	60%

Caption: n = number of participants



Caption: A1 = Child-Youth Communication and Language; A2 = Voice and Orofacial Motor Skills; A3 = Cognition and Communication in Adults and the Elderly; A4 = Audiology and Vestibular

Figure 1. Procedure for the validation of the Scripts corpus



Caption: A-C = Child-Youth Communication and Language; D-F = Voice and Orofacial Motor Skills; G-I = Cognition and Communication in Adults and the Elderly; J-L = Audiology and Vestibular

Figure 2. Final scores by area and type of script

Table 2. Corpus Reliability by area and type of scripts

Area	Type of Script	α Cronbach
Child-Youth Communication and Language	Diagnostic	0.66
	Study	0.84
	Intervention	0.59
Voice and Orofacial Motor Skills	Diagnostic	0.68
	Study	0.82
	Intervention	0.72
Cognition and Communication in Adults and the Elderly	Diagnostic	0.69
	Study	0.82
	Intervention	0.65
Audiology and Vestibular	Diagnostic	0.69
	Study	0.34
	Intervention	0.64

Table 3. Fleiss's Kappa Index by area and type of Script

Area	Type	Combined Kappa	Z	Prob > Z
Child-Youth Communication and Language	Diagnostic	0.38	25.43	0.0001
	Study	0.25	8.99	0.0001
	Intervention	0.21	13.85	0.0001
Voice and Orofacial Motor Skills	Diagnostic	0.39	24.42	0.0001
	Study	0.07	2.07	0.0193
	Intervention	0.27	19.64	0.0001
Cognition and Communication in Adults and the Elderly	Diagnostic	0.26	15.51	0.0001
	Study	0.08	2.41	0.008
	Intervention	0.17	12.2	0.0001
Audiology and Vestibular	Diagnostic	0.45	22.76	0.0001
	Study	0.22	4.27	0.0001
	Intervention	0.22	12.47	0.0001

DISCUSSION

The diversification of teaching strategies oriented to the development of complex competencies such as CR, is key in training undergraduate health students⁽²⁴⁾. To date, experiences implementing the script agreement test have been reported in careers such as medicine and nursing⁽²⁵⁾, among others. The results of this study show variability in the judgment of the experts by area (Figure 1), which is aligned with the description of the decision-making process in uncertainty scenarios⁽²⁶⁾. It is consistent with previous research experiences where divergent responses have been described as a recurring finding in expert panels' resolution of clinical scripts⁽²³⁾.

The evaluation of the CR is a complex process in pedagogical terms⁽²⁷⁾, and the scripts have proven to be a reliable and valid teaching tool when there is a rigorous design and build process⁽²⁸⁾. In this sense, it is essential to highlight that the scripts have made it possible to discriminate differential degrees of performance according to the educational level and the time of professional experience in medicine⁽²⁸⁾; hence its application for the postgraduate course in speech-language therapy represents an interesting projection that can be explored in future research.

The use of scripts implies certain challenges when thinking about their design, construction, and implementation, specifically

related to the validity of the corpora, the handling of low scores, and the association of these results with other evaluation techniques. In this sense, and for practical purposes, as reported in the methodology, all the scripts that presented incomplete answers and those that did not have dominant agreements were eliminated from the corpus, that is, a considerably low frequency considered as such by the experts⁽²³⁾. Thus, a global reliability of $\alpha = 0.67$ was obtained, which, although considered low, is similar to what was reported in a previous study in postgraduate medicine ($\alpha > 0.65$ in all items)⁽²⁹⁾.

The consistency in the experts' answers to the scripts, in general, is acceptable beyond chance (Table 3), except in study scripts in the areas of Voice and Orofacial Motor Skills, Cognition and Communication in Adults and the Elderly. This finding is consistent in relation to the studies in other disciplines where there is a coincidence in the agreement between the validators with scores that are usually higher than those observed in novices without experience, and allow discriminating the educational level and proficiency⁽³⁰⁾. In this way, it is verified that the dossier created has properties which make it functional for evaluating CR skills in Speech Pathology students, as the evidence highlights it in other areas of health⁽²⁹⁾.

A limitation of this study is the variability in what is done and the professional judgment, which is expressed as the dispersion in response trends by experts in certain areas (Figure 2D, E, H, J) and has an impact on the reliability scores (Table 2). However, the literature offers some explanations for this phenomenon, such as the tendency of validation professionals to avoid extreme responses, that is, -2 or +2 in scripts⁽¹⁶⁾. In this line, although the overall reliability score was low, it is important to highlight that the poor reliability result ($\alpha = 0.34$) was maintained in the analysis in the case of the study scripts in the Audiology and Vestibular area, which affected the final calculation on the dossier on the one hand but avoided the confirmatory bias. Yet, the discrepancy between the experts is acceptable, considering that the validated instrument does not correspond to a diagnostic instrument.

Concerning the sample size, it is important to note that the literature recommendation was considered, so although it is possible to expand the number of experts per field, this research adhered to what was suggested by previous research⁽¹⁸⁾.

As a projection, it is interesting to incorporate validation of the content of the dossier beyond the ascription to the learning results of the subject programs, as it is to add a second evaluative instance to determine the consistency of the intra-evaluator responses, increase the number of scripts per type, and increase the number of validators. Likewise, it will be essential to determine the impact and association of incorporating this strategy with respect to students' progress beyond clinical decision-making in scenarios with high uncertainty⁽³¹⁾.

CONCLUSION

The corpus created presented a low global reliability. Still, the stratification by area and type of script offers a heterogeneous panorama, where it is possible to distinguish those items where a greater or lesser correlation was obtained, depending on the clinical reasoning task presented. Additionally, the inter-evaluator agreement was acceptable for the entire dossier. Determining the psychometric properties of scripts in the context of the script agreement test is the first formal approach to generating an innovative pedagogical tool for the training and authentic evaluation of CR in undergraduate students of the Speech-Language Therapy career. It is necessary to increase research and educational innovation in this field to offer relevant training instruments which are based on available scientific evidence and have validation to improve the development, monitoring and evaluation of professional proficiency in students of health careers, in preclinical instances and/or in early formative stages, complementing the strategies, instruments and tools already established in the curricula. The strategy developed is transferable, but it must be adjusted to the curricular frameworks of each house of study in terms of content and learning outcomes.

ACKNOWLEDGEMENTS

To Doctor Rigoberto Marín for his permanent support and guidance

REFERENCES

1. Moore P, Leighton MI, Alvarado C, Bralic C. Simulated patients in health care training: the human side of simulation. *Rev Med Chil*. 2016;144(5):617-25. <http://dx.doi.org/10.4067/S0034-98872016000500010>. PMID:27552013.
2. Huhn K, Gilliland S, Black L, Wainwright S, Christensen N. Clinical reasoning in physical therapy: a concept analysis. *Phys Ther*. 2019;99(4):440-56. <http://dx.doi.org/10.1093/ptj/pzy148>. PMID:30496522.
3. Thampy H, Willert E, Ramani S. Assessing clinical reasoning: targeting the higher levels of the pyramid. *J Gen Intern Med*. 2019;34(8):1631-6. <http://dx.doi.org/10.1007/s11606-019-04953-4>. PMID:31025307.
4. Vega Rodríguez YE, Torres Rodríguez AM, del Campo Rivas MN. Análisis del rol del Fonoaudiólogo(a) en el sector salud en Chile. *Cienc Trab*. 2017;19(59):76-80. <http://dx.doi.org/10.4067/S0718-24492017000200076>.
5. del Campo M, Silva-Ríos A, Valdés J. Perspectivas y desafíos de los fonoaudiólogos en la actividad académica en Chile: una descripción preliminar. *Rev Chil Fonoaudiol*. 2019;18:1-10. <http://dx.doi.org/10.5354/0719-4692.2019.55330>.
6. Linn A, Khaw C, Kildea H, Tonkin A. Clinical reasoning: a guide to improving teaching and practice. *Aust Fam Physician*. 2012;41(1-2):18-20. PMID:22276278.
7. Delany C, Golding C. Teaching clinical reasoning by making thinking visible: an action research project with allied health clinical educators. *BMC Med Educ*. 2014;14(1):20. <http://dx.doi.org/10.1186/1472-6920-14-20>. PMID:24479414.
8. Yazdani S, Hosseinzadeh M, Hosseini F. Models of clinical reasoning with a focus on general practice: a critical review. *J Adv Med Educ Prof*. 2017;5(4):177-84. PMID:28979912.
9. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med*. 2000;75(2):182-90. <http://dx.doi.org/10.1097/00001888-200002000-00020>. PMID:10693854.
10. Fayol M, Monteil JM. The notion of script: from general to developmental and social psychology. *Cah Psychol Cogn Psychol Cogn*. 1988;8(4):335-61.
11. Delavari S, Monajemi A, Baradaran HR, Myint PK, Yaghmaei M, Soltani Arabshahi SK. How to develop clinical reasoning in medical students and interns based on illness script theory: an experimental study. *Med J Islam Repub Iran*. 2020;34:9. <http://dx.doi.org/10.47176/mjiri.34.9>. PMID:32284933.
12. Iravani K, Amini M, Doostkam A, Dehbozorgian M. The validity and reliability of script concordance test in otolaryngology residency training. *J Adv Med Educ Prof*. 2016;4(2):93-6. PMID:27104204.
13. Nazim SM, Talati JJ, Pinjani S, Biyabani SR, Ather MH, Norcini JJ. Assessing clinical reasoning skills using Script Concordance Test (SCT) and extended matching questions (EMQs): a pilot for urology trainees. *J Adv Med Educ Prof*. 2019;7(1):7-13. PMID:30697543.
14. Wan SH. Using the script concordance test to assess clinical reasoning skills in undergraduate and postgraduate medicine. *Hong Kong Med J*. 2015;21(5):455-61. <http://dx.doi.org/10.12809/hkmj154572>. PMID:26314569.
15. Custers EJFM. The script concordance test: an adequate tool to assess clinical reasoning? *Perspect Med Educ*. 2018;7(3):145-6. <http://dx.doi.org/10.1007/S40037-018-0437-6>. PMID:29904901.
16. See KC, Tan KL, Lim TK. The script concordance test for clinical reasoning: re-examining its utility and potential weakness. *Med Educ*. 2014;48(11):1069-77. <http://dx.doi.org/10.1111/medu.12514>. PMID:25307634.
17. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ*. 2013;47(12):1175-83. <http://dx.doi.org/10.1111/medu.12283>. PMID:24206151.
18. Arceo M, Durante E. Desarrollo y evaluación de los scripts durante la formación profesional. *Rev Hosp Ital B Aires [Internet]*. 2013 [citado 2022 abril 28];33(4):144-52. Disponible en: https://www.hospitalitaliano.org.ar/multimedia/archivos/noticias_attachs/47/documentos/16200_144-152-HI-4-9-Ed%20Medica-Arceo-E.pdf

19. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak*. 2008;8(1):18. <http://dx.doi.org/10.1186/1472-6947-8-18>. PMID:18460199.
20. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ*. 2012;46(6):552-63. <http://dx.doi.org/10.1111/j.1365-2923.2011.04211.x>. PMID:22626047.
21. Manterola C, Grande L, Otzen T, García N, Salazar P, Quiroz G. Confiabilidad, precisión o reproducibilidad de las mediciones: métodos de valoración, utilidad y aplicaciones en la práctica clínica. *Rev Chilena Infectol*. 2018;35(6):680-8. <http://dx.doi.org/10.4067/S0716-10182018000600680>. PMID:31095189.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74. <http://dx.doi.org/10.2307/2529310>. PMID:843571.
23. Gagnon R, Lubarsky S, Lambert C, Charlin B. Optimization of answer keys for script concordance testing: should we exclude deviant panelists, deviant responses, or neither? *Adv Health Sci Educ Theory Pract*. 2011;16(5):601-8. <http://dx.doi.org/10.1007/s10459-011-9279-2>. PMID:21286807.
24. Wan MSH, Tor E, Hudson JN. Construct validity of script concordance testing: progression of scores from novices to experienced clinicians. *Int J Med Educ*. 2019;10:174-9. <http://dx.doi.org/10.5116/ijme.5d76.1eee>. PMID:31562807.
25. Deschênes MF, Charlin B, Gagnon R, Goudreau J. Use of a script concordance test to assess development of clinical reasoning in nursing students. *J Nurs Educ*. 2011;50(7):381-7. <http://dx.doi.org/10.3928/01484834-20110331-03>. PMID:21449528.
26. Gawad N, Wood TJ, Cowley L, Raiche I. How do cognitive processes influence script concordance test responses? *Med Educ*. 2021;55(3):354-64. <http://dx.doi.org/10.1111/medu.14416>. PMID:33185303.
27. Faucher C, Dufour-Guindon MP, Lapointe G, Gagnon R, Charlin B. Assessing clinical reasoning in optometry using the script concordance test. *Clin Exp Optom*. 2016;99(3):280-6. <http://dx.doi.org/10.1111/cxo.12354>. PMID:27087346.
28. Karila L, François H, Monnet X, Noel N, Roupret M, Gajdos V, et al. The script concordance test: a multimodal teaching tool. *Rev Med Interne*. 2018;39(7):566-73. <http://dx.doi.org/10.1016/j.revmed.2017.12.011>. PMID:29576195.
29. Subra J, Chicoulaa B, Stillmunkès A, Mesthé P, Oustric S, Rougé Bugat ME. Reliability and validity of the script concordance test for postgraduate students of general practice. *Eur J Gen Pract*. 2017;23(1):208-14. <http://dx.doi.org/10.1080/13814788.2017.1358709>. PMID:28819998.
30. Kow N, Walters MD, Karram MM, Sarsotti CJ, Jelovsek JE. Assessing intraoperative judgment using script concordance testing through the gynecology continuum of practice. *Med Teach*. 2014;36(8):724-9. <http://dx.doi.org/10.3109/0142159X.2014.910297>. PMID:24819908.
31. Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM. Script concordance testing: a review of published validity evidence. *Med Educ*. 2011;45(4):329-38. <http://dx.doi.org/10.1111/j.1365-2923.2010.03863.x>. PMID:21401680.

Author contributions

APSR and MNCR participated in the study design, setup, data analysis and interpretation, article writing, review, and critical reading; PKKU and VACS contributed to the setup, coordination, and data collection, as well as background tabulation, writing, and revision of the manuscript.