Triple Jump. An Alternative to be Consider For Assessing Compliance with Goals in Graduate Programs

Salto triple. Una alternativa de evaluación del cumplimiento de objetivos en programas de postgrado

Carlos Manterola[1,2]; Ricardo Cartes-Velásquez[3] & Tamara Otzen[4]

SUMMARY: Ensuring quality in graduate teaching is a guarantee of the competitiveness demonstrated by institutions. The aim of this study was to determine the internal consistency and reliability of the Triple Jump (TJ) instrument applied to Master's and Doctorate students in Medical Sciences to assess compliance of goals. Reliability study conducted at Universidad de La Frontera, Temuco, Chile. TJ was applied by 3 evaluators who performed 2 evaluations, with structured assessment guidelines in three domains (cognitive, attitudinal and procedural), on a Likert scale with scores from 0 (not achieved) to 4 (achieved). The total point score was transformed into a final grade. The evaluated subjects were characterized, and internal consistency and interobserver reliability were determined. Fifty-one Health care professionals enrolled in Medical Sciences graduate programs at the Universidad de La Frontera between 2005-2016 (35 from the Master's program and 16 from the Doctorate in Medical Sciences). In study subjects had a median age of 29 years, 6 years of professional experience and 3 years as specialists. 66.7 % were male. The internal consistency of the responses to the assessment was characterized by a Cronbach's between 0.734 and 0.938. Interobserver reliability of the instrument had values between 0.86 and 1.0. The levels of internal consistency and reliability were higher than previously reported; therefore, the test is reliable and exhibits solid internal consistency.

KEY WORDS: Graduate education; Assessment; Triple jump; Practice-Based Learning; Internal consistency; Reliability.

INTRODUCTION

Ensuring quality in graduate teaching is a guarantee of the competitiveness demonstrated by institutions. It indicates the ability to fulfill the educational objectives drawn up as part of the organization's academic purposes, and it is the surest way to satisfy both students' aspirations and professors' expectations (Rosselot, 1999; Rosselot *et al.*, 2002).

Devising a testing procedure in a graduate program is a great challenge for the faculty, since instruments of quality and suitability are needed (Polit & Hungler, 2000; Sacristán, 2015).

The triple jump (TJ) is a three-step oral evaluation promoted by McMaster University. It has a reported interobserver reliability between 0.5 and 0.8 (Chong *et al.*, 1984) and between 0.75 and 0.87 (Chapman *et al.*, 1993).

The TJ gives both the student and teacher the opportunity to simulate a real situation that evaluates the student's understanding of concepts, previous knowledge, skills in self-directed learning, self-assessment and problem-solving abilities. Since it features flexibility and adaptability to different situations and the students' levels of experience, it enables observation of the skill levels in collecting and ordering the data obtained (Chapman *et al.*; Navazesh *et al.*, 2013; Moore & Vaughan, 2015). It also allows for personal interaction between students, the evaluation among peers and the self-evaluation of each student.

Problem-based or process-focused learning as a pedagogical approach in health care education is an option that has gradually come to predominate over the years (Dueñas, 2001; Bordas Alsina & Cabrera Rodríguez, 2001; Castro *et al.*, 2003; Matthes *et al.*, 2008; Navarro *et al.*, 2010;

[1] Department of Surgery, Universidad de La Frontera, Chile.
[2] Center for Morphological and Surgical Studies (CEMyQ), Universidad de La Frontera, Chile.
[3] School of Dentistry, Universidad de Concepción, Chile.
[4] Faculty of Health Sciences, Universidad de Tarapacá, Chile.

Allen *et al*., 2016). However, most experience in the application of these strategies has concentrated on undergraduate education, and since the outcomes have been satisfactory, it seemed to us that graduate education could be an interesting scenario to investigate the behavior of such tools at this level of university education.

The TJ evaluation is part of the assessment process in the Master's and Doctorate programs in Medical Sciences offered by the Universidad de La Frontera. Nevertheless, it does not have a psychometric evaluation of the instrument applied in this type of evaluation. As a result, it is necessary to assess the validity and reliability of our experience of evaluation using the TJ (Moore & Vaughan).

The aim of this study was to determine the internal consistency and reliability of the TJ instrument applied to Master's and Doctorate students in Medical Sciences to assess compliance of goals.

MATERIAL AND METHOD

Study design: Reliability study.

Setting: Master's and Doctorate Programs in Medical Sciences, Universidad de La Frontera, Temuco, Chile.

Participants: Health care professionals from different institutions and various disciplines enrolled in the Master's or Doctorate program in Medical Sciences at the Universidad de La Frontera 2005-2016 were included. There were no exclusion criteria.

Sampling: A non-probability sampling of consecutive cases was applied.

Variables: Biodemographic variables of the evaluated subjects, internal consistency and interobserver reliability of the instrument applied to conduct the TJ.

TJ stages: There are three stages: i) definition of the problem, ii) search for information and study, and iii) formulation and final synthesis of the problem (Painvin *et al*., 1979; Painvin, 1981; Powles *et al*., 1981; Matthes *et al*.; Navarro *et al*.). The instrument is applied at stages one and three of the process (Fig. 1 and Table I).

In the first stage, called "Problem definition", in approximately 15 minutes the student and examiner(s) establish an exploratory conversation based on a scenario/problem, where the student recognizes the problem areas, reviews the knowledge acquired already and identifies the knowledge to be acquired. Hypotheses are posed and the second stage is planned (Chapman *et al*.).

In the second stage, called "Information search and study", which lasts two to 24 hours (depending on the evaluation objectives drawn up, which in our case was 24 hours), the student performs an independent study using different resources (human, bibliographic, scientific, audiovisual, institutional, services, etc.). The information search is meant to respond to the questions the student has formulated during the first stage in the form of learning or content goals (Chapman *et al*.).
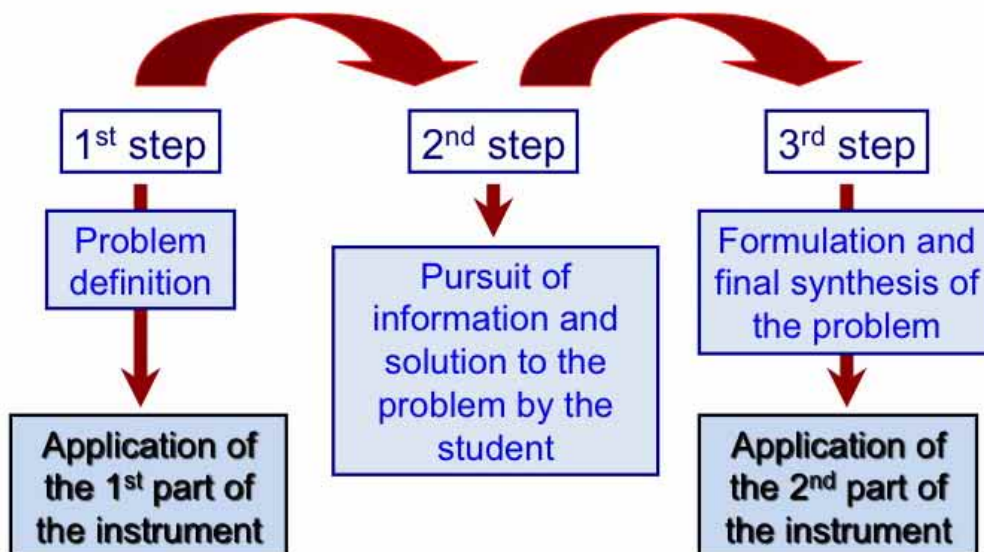

Fig. 1. Description of the evaluation process. TJ stages.

Items

STAGE 1:

1. Collecting information (Soliciting or investigating topics)

0	1	2	3	4
Does not know where to start				Asks relevant questions with ease

2. Organization

0	1	2	3	4
Acquisition of information was disorganized				Acquired organized and logical information

3. Interpreting and positing a hypothesis

0	1	2	3	4
Unable to suggest a logical explanation with data provided				Able to give a logical explanation with data provided

4. Relevance to the problem, identification of topics

0	1	2	3	4
The proposed topics to explore were irrelevant to the proposed problem				Was able to identify topics or materials of interest

5. Establishing exploration priorities

0	1	2	3	4
Enumerates without clearly identifying an order of priorities				Clearly establishes the priority elements of the problem

6. Applying previous knowledge

0	1	2	3	4
Shows little basic knowledge in relation to the problem				Shows knowledge and understanding of principles and concepts regarding the problem

Items

STAGE 3:

7. Searching for information

0	1	2	3	4
Poorly organized study period poorly utilized (or chosen) resources				Well planned study period, adequate resources optimal synthesis of the data

8. Using the information

0	1	2	3	4
Uses the resources to resolve the case in an inconsistent way				Efficiently uses the information to resolve the case

9. Exploring topics identified in stage 1 (contents)

0	1	2	3	4
Does not explore the topics, no consistency with stage 1. Does not try to explore other ideas				Explored the topics, and the additional ideas are appropriate and consistent with stage 1

10. Integrating knowledge

0	1	2	3	4
Shows no integration of knowledge of the different areas covered in-depth in the program				Adequately integrates knowledge acquired in the program

11. Managing the discussion

0	1	2	3	4
Shows no command of the topic, does not respond to questions				Shows command of the topic, responds to all questions adequately

12. Organizing the presentation

0	1	2	3	4
Does not adequately allocate time, does not highlight relevant aspects				Optimal allocation of time, highlights relevant aspects

13. Explaining and/or analyzing the problem

0	1	2	3	4
Unable to explain the problem and reach conclusions				Able to demonstrate that in large part the problem can be explained

14. Identifying alternative problems

0	1	2	3	4
Unable to identify elements that must be explored later				Identifies areas of relevance that must be explored later

15. Self-evaluation

0	1	2	3	4
Unable and unwilling to evaluate his own performance				Clear and systematic evaluation of his performance Adds ideas of how to improve

Table I. Tj Assessment Guideline

In the third stage, called "Formulation and final synthesis of the problem", which takes 15 to 30 minutes, the student returns for a brief presentation about the process undertaken in stage 2, his conceptual synthesis, fulfillment of goals (contents), corroboration of the hypothesis and conclusions. The student and the examiners evaluate the performance, including the resources utilized and the efficiency of their use (Chapman *et al.*).

Instrument Features: A structured instrument is applied with criteria and attitudinal, procedural and cognitive value judgments on a Likert scale from 0 (not achieved) to 4 (achieved), that is generated from discussions of the multidisciplinary faculty that comprise the two programs.

The criteria to evaluate in the first stage are: Problem identification, acquisition of information and interpretation (question formulation), application of previous knowledge according to the level of advancement in the program, posing of goals and priorities (or activities), confrontation of difficulties generated by the problem. Several cases problem from different types of scenarios were used, including some of therapy, diagnosis and prognosis. The problem cases used different type of scenarios, including: therapy, prognosis, and diagnosis; in all cases was used the same instrument. All of them were evaluated using the instrument shown in Table I.

In the second stage, the student looks for and selects the information to solve the problem posed in the previous stage. It may resort to different sources of information, including: people (quality of experts on a subject), audio-visual media, Internet, traditional texts, and references available in the several databases for biomedical literature. To do this, he or she must use MeSH terms, free words and search strategies, after which he must evaluate these studies critically, applying user guides (critical reading guides) to ultimately summarize the data obtained.

The criteria to evaluate in the third stage are: information search (organization and selection of the resources used), recognition of what is relevant (ability to recognize and synthesize it), skill in using the data obtained, synthesis of the goals or contents raised in the first stage, knowledge integration, contribution to solving the problem, organization of the presentation, quality of the visual material used in the presentation, management of the discussion, and self-evaluation (Table I).

The assessment of each of these items generates a score (from 0 to 4), the sum of which is transformed into a final grade on a scale from 1.0 to 7.0, with a pass being a minimum of 5.0.

Maneuver: The student must solve a specific problem, where he will apply the knowledge and skills acquired in the program to provide a concrete solution to the issue. We seek a connection that relates the areas of conceptual bases of clinical practice, clinical epidemiology, biostatistics, critical literature analysis, research designs, bioethics and the student's own professional experiences.

This integrative evaluation is applied at the end of the first and the second semesters of each program by 3 evaluators (all of them are professors of both programs and with PhD degree).

Statistical tools: Descriptive statistics was applied (calculation of percentages, measures of central tendency and extreme values) to characterize the subjects. Then, calculating Cronbach's determined the internal consistency of the items ? and finally, the interobserver reliability was considered by applying the intraclass correlation coefficient.

Ethical principles: The identity of the participants (students and evaluators) was kept hidden by encoding the data.

RESULTS

In the study period, a total of 51 students from the two programs (35 in the Master's and 16 to in the Doctorate program) were evaluated using the TJ. With a median age of 29 years (25 to 51 years), and a median of professional experience and as specialists of 6 years (2 to 28 years) and 3 years (1 to 20 years), respectively. 66.7 % were male.

Their professional activities were: physicians (25) specializing in general, plastic or vascular surgery, obstetrics and gynecology, urology; dentists (18) specializing in maxillofacial trauma surgery, periodontics, implantology and temporomandibular disorders; physiotherapists (5), nutritionists (1), psychologists (1) and medical technologists (1).

Internal consistency of the responses to the evaluation had a Cronbach's between 0.734 and 0.938 according to the stage of the evaluative process and when it was applied (1st or 2nd semester). (Table II).

Interobserver reliability had an intraclass correlation coefficient between 0.86 and 1.0 for parts 1 and 2 of the instrument in the 1st semester application (throughout the years studied), and between 0.92 and 1.00 for parts 1 and 2 of the instrument in the 2nd semester application (throughout the years studied). (Table III).

Table II. Internal consistency of the response to the evaluation process.

Process	Steps	Items	Nº Evaluations	Cronbach's
1st semester	1	1-6	52	0.828
1st semester	2	7-15	52	0.734
2nd semester	1	1-6	52	0.843
2nd semester	2	7-15	52	0.938

Table III. Interobserver reliability of the 1st and 2nd evaluation.

Year	Semester	Part 1	Part 2
2004	1st	0.98	0.99
	2nd	0.93	0.99
2005	1st	0.91	0.95
	2nd	0.98	0.97
2006	1st	0.95	0.98
	2nd	0.97	0.99
2007	1st	0.99	0.92
	2nd	0.98	0.94
2008	1st	0.86	0.99
	2nd	0.90	0.98
2009	1st	NR	NR
	2nd	0.97	0.99
2010	1st	0.99	0.99
	2nd	0.98	0.99
2011	1st	0.98	0.99
	2nd	0.91	0.97
2012	1st	0.99	0.97
	2nd	0.98	0.99
2013	1st	0.99	0.97
	2nd	0.95	0.97
2014	1st	0.99	0.98
	2nd	0.97	0.84
2015	1st	0.96	0.94
	2nd	0.98	0.99

NR: Not realized

DISCUSSION

It seems to us of interest to indicate that this type of evaluation is a theoretical-practical activity, integrative and formative in nature. This affords the student the opportunity to understand his level of progress in the program better in terms of the cognitive, attitudinal and procedural aspects and for the academics to participate in an integrative evaluation. The TJ even allows the student to be approached more closely and observed trying to solve the problem posed at different points in time (Chapman *et al*.; Navazesh *et al*.; Moore & Vaughan), applying previous knowledge as well as knowledge obtained in the graduate program.

Referring to a conceptual approach of the process, we can highlight that research conducted in education (Rinaudo, 1995; Atienza Cerezo & López Ferrero, 1995; Tapia el al., 2003), has reported adults having difficulty performing cognitive operations frequently related to areas of very specialized knowledge; we are referring in this context to the Master's degree. As academics in graduate programs strive for students to reach a global and complex view of reality, this means offering them support to try and bring clarity of judgment and ideas, where the student, through the search for indications, necessarily increases his knowledge for making decisions, developing competencies in sifting through information, managing to synthesize and obtain analytical and interpretative frameworks of information without leaving the social and bioethical domains aside.

The operating method of this assessment experience has a high anticipatory and explanatory potential in decision-making from the simulation of a real situation. Here the human criterion is fundamental, because the information is organized systematically and in favor of the subject (MINSAL, 1997), simplifying the complex without losing its value (Gómez García, 2003).

With respect to the psychometric analysis of the instrument used to evaluate the TJ, we can say that relevance, thoroughness and exclusivity of the items support the substantial and significant aspects of the evaluated goals, including all the possible contents, and these do not overlap. The homogeneity presented of the items in each dimension advocates a solid internal structure, since it has to do with the degree of equivalence in relation to the measurement of the dimension, which represents an adequate construct validity (Streiner & Norman, 2006). Although the second section evaluated in the second semester displayed a certain redundancy with a Cronbach's $a > 0.9$, the rest of the values were at an acceptable value of 0.7 to 0.9 (Campo-Arias & Oviedo, 2008).

In relation to the level of reliability of the domains, it is within the permissible limits to describe the test as adequate and reproducible, since coefficients that range between 0.80 and 0.85 are accepted (Salas Perea, 1998). These values are higher than those reported with similar instruments for the evaluation of undergraduate students in the area of health care (Asuncion *et al*., 2013; Hurley *et al*., 2015).

With respect to the "advantages of the evaluative process" identified by those involved, they emphasized the following:

a) The greatest challenge was the information management, being favorable because it facilitates actions towards improvement in terms of effectiveness, efficiency and relevance.
b) The student is able to recognize the degree of certainty and uncertainty in which he moves for making decisions and with these dimensions apply his criteria.
c) They live the experience of applying the acquired tools, which makes their learning expectations more satisfying.
d) There is autonomy in the decisions; the student himself poses the work plan with goals to meet.
e) It gives the student a unique view of a whole, i.e., from the multiplicity of areas and subjects, he can relate, prioritize and focus.
f) It allows the development of such aspects as ability to synthesize, language and use of visual support resources for an oral presentation.
g) There is an integration of the academic team with varying ranks and types of experience, which lends a multidimensional view to the problem.

With respect to the disadvantages, the following can be mentioned:

a) Inherent resistance that such an action - the application of an integrating evaluative process-provokes.

b) As academics we tend to express our own experience in resolving situations, we need to "contain ourselves ". The evaluators must take care with their advice, i.e., orient the student without giving the answer, guide so that he looks for the answer himself.

c) The time to prepare the cases must be added to the time involved in fulfilling the stages. The academics must allocate long periods in preparing the cases and leading them to discussion; it is necessary to verify the clarity of the writing in line with the goals to be evaluated. In the third stage, the time allocated for each student is at least 20 minutes for the presentation, with 15 minutes added for discussion.

There is evidence indicating that performance of this assessment instruments is case specific and some evidence suggesting that correlations across cases are small (Epstein, 2007; Norcini & McKinley, 2007; Setyonugroho *et al.*, 2015), however our experience using different problem cases of different types of scenarios (therapy, diagnosis and prognosis) has been satisfactory.

Our study has several limitations. First, it was conducted in two programs at the same institution. Other institutions with different student populations, curriculum designs or tutor training and quality may yield different results. Second, we used an in-house instrument for assessing TJ performance. Although the domains assessed by the tool tended to be commonly used, the detailed behavioral indicators or descriptors for each domain and the rating scale used may differ between institutions. Third, our findings did not provide conclusive support for interobserver reliability because it seems to us a preliminary analysis that suggests indications; therefore, we can say that it shows consistency and stability to certain degree (for a better approach, a greater number of evaluations and new tests are necessary). Finally, the design of using this type of evaluation prevented us from examining other factors, such as the quality and complexity of scenarios, and dynamic interactions between group members as suggested by other researchers (Norman & Schmidt, 2000; Walker & Leary, 2009), which may affect the reliability and validity of the assessment scenario.

CONCLUSIONS

We think TJ is an assessment is useful in graduate education (more than traditional comprehensive exams), because is more practical and applied, and on the other hand it is more clinically relevant for students who are currently in or will be working in the healthcare field.

Postgraduate programs need better assessments of student competency and TJ is an example of quality assessment to consider by graduate programs.

Internal consistency and reliability of the TJ instrument was determined to evaluate significant elements of a graduate student's learning in medical sciences.

The instrument used for this type of evaluation presents internal consistency and interobserver reliability in this report higher than those previously published.

FINANCIAL DISCLOSURE

MANTEROLA, C.; CARTES-VELÁSQUEZ, R. & OTZEN, T. Salto triple. Una alternativa de evaluación del cumplimiento de objetivos en programas de postgrado. *Int. J. Morphol., 36(1)*:373-379, 2018.

RESUMEN: Garantizar la calidad en la enseñanza de posgrado es una garantía de la competitividad demostrada por las instituciones. El objetivo de este estudio fue determinar la consistencia interna y confiabilidad interobservador del instrumento salto triple (ST), aplicado a estudiantes de Magíster y Doctorado en Ciencias Médicas, para evaluar cumplimiento de objetivos de aprendizaje. Estudio de confiabilidad realizado en la Universidad de La Frontera, Temuco, Chile. El ST, fue aplicado por 3 evaluadores que realizaron 2 evaluaciones, utilizando guías de evaluación estructuradas en tres dimensiones (cognitivo, actitudinal y de procedimiento); en una escala tipo Likert con puntajes de 0 (objetivo no logrado) a 4 (objetivo plenamente logrado). El puntaje total se transformó posteriormente en una calificación final. Se caracterizó a los sujetos evaluados y se determinó consistencia interna y confiabilidad interobservador. Para ello, se enrolaron 51 profesionales de la salud matriculados en los programas de postgrado en Ciencias Médicas de la Universidad de La Frontera entre 2005 y 2016 (35 del programa de Magíster y 16 del Doctorado en Ciencias Médicas). Los sujetos a estudio, tenían una mediana de edad de 29 años, 66,7 % eran hombres, tenían 6 años de experiencia profesional y 3 años como especialistas. La consistencia interna de las respuestas a la evaluación se caracterizó por un a de Cronbach entre 0,734 y 0,938. La confiabilidad interobservador del instrumento tenía valores entre 0,86 y 1,0. Los niveles de consistencia interna y confiabilidad fueron más altos que los reportados previamente; por ende, nos parece que esta forma de evaluar objetivos en este tipo de estudiantes, tiene una consistencia interna sólida y es confiable.

PALABRAS CLAVE: Educación de postgrado; Evaluación; Salto triple; Enseñanza basada en solución de problemas; Consistencia interna; Confiabilidad.

REFERENCES

Allen, S. J.; King, M. A. & Yeadon, M. R. Optimisation of phase ratio in the triple jump using computer simulation. *Hum. Mov. Sci., 46*:167-76, 2016.

Asuncion, A. M.; Cagande, C.; Schlagle, S.; McCarty, B.; Hunter, K.; Milcarek, B.; Staman, G.; Da Silva, S.; Fisher, D. & Graessle, W. A Curriculum to improve Residents' End-of-Life communication and pain management skills during pediatrics intensive care rotation: Pilot study. *J. Grad. Med. Educ., 5(3)*:510-3, 2013.

Atienza Cerezo, E. & López Ferrero, C. El contexto en el discurso académico: Su influencia en la presentación y desarrollo de la información. *Rev. Pedag., (10)*:123-9, 1995.

Bordas Alsina, M. I. & Cabrera Rodríguez, F. A. Estrategias de evaluación de los aprendizajes centradas en el proceso. *Rev. Esp. Pedag., 59(218)*:25-48, 2001.

Campo-Arias, A. & Oviedo, H. C. Propiedades psicométricas de una escala: La consistencia interna. *Rev. Salud Publ., 10(5)*:831-9, 2008.

Castro, A. M.; Reyes, R. M. A.; Coria, C. R.; García, Y. Y. & Gutiérrez, Q. M. Experiencia en la aplicación del método del aprendizaje basado en problemas en una asignatura de libre elección. *Rev. Fac. Med. U. N. A. M., 46(6)*:246-50, 2003.

Chapman, J. A.; Westmorland, M. G.; Norman, G. R.; Durrell, K. & Hall, A. The structured oral self-directed learning evaluation: one method of evaluating the clinical reasoning skills of occupational therapy and physiotherapy students. *Med. Teach., 15(2-3)*:223-36, 1993.

Chong, J. P.; Neufeld, V.; Oates, M. J. & Secord, M. The selection of priority problems and conditions: an innovative approach to curriculum design in medical education. *Res. Med. Educ., 23*:17-25, 1984.

Dueñas, V. H. El aprendizaje basado en problemas como enfoque pedagógico en la educación en salud. *Colomb. Med., 32(4)*:189-96, 2001.

Epstein, R. M. Assessment in medical education. *N. Engl. J. Med., 356(4)*:387-96, 2007.

Gómez García, P. *La Antropología Compleja.* In: Morin, E. (Ed.). Homo Complexus. Granada, Editorial Universidad de Granada, 2003.

Hurley, K. F.; Giffin, N. A.; Stewart, S. A. & Bullock, G. B. Probing the effect of OSCE checklist length on inter-observer reliability and observer accuracy. *Med. Educ. Online, 20*:10.3402/meo.v20.29242, 2015.

Matthes, J.; Look, A.; Hahne, A. K.; Tekian, A. & Herzig, S. The semi-structured triple jump--a new assessment tool reflects qualifications of tutors in a PBL course on basic pharmacology. Naunyn Schmiedebergs *Arch. Pharmacol., 377(1)*:55-63, 2008.

Ministerio de Salud de Chile (MINSAL). *Cuidado de Salud Basado en Evidencia: Una Salud Efectiva al Servicio de las Personas, en Diseño e Implementación de las Prioridades en Salud. La Reforma Programática Chilena.* Santiago de Chile, Ministerio de Salud de Chile (MINSAL), 1997.

Moore, K. & Vaughan, B. Assessment of Australian osteopathic learners' clinical competence during workplace learning. *Int. J. Osteopath. Med., 19*:50-60, 2015.

Navarro, H. N.; Illescas, P. M. & Cabezas, G. M. Salto triple grupal: una estrategia evaluativa del proceso de aprendizaje basado en problemas. *Rev. Educ. Cienc. Salud, 7(2)*:112-8, 2010.

Navazesh, M.; Rich, S. K.; Chopiuk, N. B. & Keim, R. G. Triple jump examinations for dental student assessment. *J. Dent. Educ., 77(10)*:1315-20, 2013.

Norcini, J. J. & McKinley, D. W. Assessment methods in medical education. *Teach. Teac. Educ., 23(3)*:239-50, 2007.

Norman, G. R. & Schmidt, H. G. Effectiveness of problem-based learning curricula: theory, practice and paper darts. *Med. Educ., 34(9)*:721-8, 2000.

Painvin, C. The triple jump: a structured technic for the evaluation of problem solving and self-learning aptitude. *Union Med. Can., 110(6)*:504-7, 1981.

Painvin, C.; Neufeld, V.; Norman, G.; Walker, I. & Whelan, G. The "triple jump" exercise--a structured measure of problem solving and self directed learning. *Annu. Conf. Res. Med. Educ., 18*:73-7, 1979.

Polit, D. & Hungler, B. (Eds.). *Investigación Científica en Ciencias de la Salud.* 6th ed. Ciudad de México, McGraw-Hill Interamericana, 2000.

Powles, A. C.; Wintrip, N.; Neufeld, V. R.; Wakefield, J. G.; Coates, G. & Burrows, J. The "triple-jump" exercise--further studies of an evaluative technique. *Annu. Conf. Res. Med. Educ., 20*:74-9, 1981.

Rinaudo, M. C. (Ed.). *Estudiar y Aprender. Investigaciones sobre la Comprensión y Aprendizaje de Textos.* Córdoba, Universidad de Córdoba, 1995.

Rosselot, E.; Norero, C.; Hanne, C. & Mateluna, E. Professional formation and its accreditation in medicine. A paradigm to sustain public confidence. *Rev. Med. Chile, 130(5)*:585-9, 2002.

Rosselot, J. E. Professional quality assurance. A new ethic frame for the practice of medicine. *Rev. Med. Chile, 127(11)*:1375-83, 1999.

Sacristán, J. A. Clinical research and medical care: towards effective and complete integration. *BMC Med. Res. Methodol., 15*:4, 2015.

Salas Perea, R. S. (Ed.). *La Evaluación en la Educación Superior Contemporánea*. La Paz, Biblioteca de Medicina XXIV, UMSA, 1998.

Setyonugroho, W.; Kennedy, K. M. & Kropmans, T. J. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. *Patient Educ. Couns., 2015.* [Epub ahead of print]

Streiner, D. L. & Norman, G. R. "Precision" and "accuracy": two terms that are neither. *J. Clin. Epidemiol., 59(4)*:327-30, 2006.

Tapia, M.; Burdiles, G. & Arancibia, B. Aplicación de una pauta diseñada para evaluar informes académicos universitarios. *Rev. Signos, 36(54)*:249-57, 2003.

Walker, A. & Leary, H. A problem-based learning meta-analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdiscip. J. Problem-Based Learn., 3(1)*:12-43, 2009.

Corresponding author:
Dr. Carlos Manterola
Department of Surgery
Universidad de La Frontera
Manuel Montt 112, Office 408
Temuco - CHILE

E-mail: carlos.manterola@ufrontera.cl