





Progress Test in Medical School: a Systematic Review of the Literature

Teste de Progresso na Escola Médica: uma Revisão Sistemática acerca da Literatura

Ademir Garcia Reberti^I 

Nayme Hechem Monfredini^I 

Olavo Franco Ferreira Filho^{II} 

Dalton Francisco de Andrade^{III} 

Carlos Eduardo Andrade Pinheiro^{III} 

Jean Carl Silva^I 

ABSTRACT

Progress Test is an objective assessment, consisting of 60 to 150 multiple-choice questions, designed to promote an assessment of the cognitive skills expected at the end of undergraduate school. This test is applied to all students on the same day, so that it is possible to compare the results between grades and analyze the development of knowledge performance throughout the course. This study aimed to carry out a systematic and literary review about Progress Test in medical schools in Brazil and around the world, understanding the benefits of its implementation for the development of learning for the student, the teacher and the institution. The study was carried out from July 2018 to April 2019, which addressed articles published from January 2002 to March 2019. The keywords used were: "Progress Test in Medical Schools" and "Item Response Theory in Medicine" in the PubMed, Scielo, and Lilacs platforms. There was no language limitation in article selection, but the research was carried out in English. A total of 192,026 articles were identified, and after applying advanced search filters, 11 articles were included in the study. The Progress Test (PTMed) has been applied in medical schools, either alone or in groups of partner schools, since the late 1990s. The test results build the students' performance curves, which allow us to identify weaknesses and strengths of the students in the several areas of knowledge related to the course. The Progress Test is not an exclusive instrument for assessing student performance, but it is also important as an assessment tool for academic management use and thus, it is crucial that institutions take an active role in the preparation and analysis of this assessment data. Assessments designed to test clinical competence in medical students need to be valid and reliable. For the evaluative method to be valid it is necessary that the subject be extensively reviewed and studied, aiming at improvements and adjustments in test performance.

KEY-WORDS

- Progress Test.
- Statistics.
- Medical Education.
- Educational Evaluation.

^IUniversidade da Região de Joinville, Joinville, Santa Catarina, Brasil.

^{II}Universidade Estadual de Londrina, Londrina, Paraná, Brasil.

^{III}Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil.

PALAVRAS-CHAVE

- Teste de Progresso.
- Estatística.
- Educação Médica.
- Avaliação Educacional.

RESUMO

O Teste de Progresso é uma avaliação objetiva, estruturada com 60 a 150 questões de múltipla escolha, elaborada com o objetivo de promover uma avaliação das competências cognitivas esperadas no final do curso de graduação. Esse teste é aplicado a todos os discentes, no mesmo dia, de modo que seja possível comparar os resultados entre as séries e analisar a performance evolutiva do conhecimento no decorrer do curso. Este trabalho teve como objetivo realizar uma revisão sistemática e literária acerca do Teste de Progresso nas escolas médicas no Brasil e no mundo, compreendendo os benefícios de sua implantação para o desenvolvimento do aprendizado, tanto para o aluno quanto para o docente e para a instituição. A pesquisa foi realizada no período de julho de 2018 a abril de 2019 e abordou artigos publicados no período de janeiro de 2002 a março de 2019. Utilizaram-se os descritores “Teste de Progresso nas escolas médicas” e “Teoria de Resposta ao Item em Medicina” nas plataformas PubMed, SciELO e Lilacs. Não houve limitação de idioma na seleção dos artigos, porém a pesquisa foi realizada em inglês. Foram encontrados 192.026 artigos, e, após a aplicação de filtros de busca avançada, incluíram-se 11 artigos no estudo. O Teste de Progresso vem sendo aplicado nas escolas médicas, de forma isolada ou em grupos de escolas parceiras, desde o final da década de 1990. Os resultados do teste constroem curvas de desempenho dos acadêmicos, o que permite identificar fragilidades e qualidades dos estudantes nas diversas áreas do conhecimento relacionadas ao curso. O Teste de Progresso não é um instrumento exclusivo de avaliação do desempenho dos estudantes, mas assume também um significado como ferramenta avaliativa para uso na gestão acadêmica, e para isso é fundamental que as instituições assumam papel ativo na elaboração e na análise dos dados dessa avaliação. Os exames desenvolvidos para testar a competência clínica em estudantes de Medicina necessitam ser válidos e confiáveis. Para que o método avaliativo seja válido, é necessário que o assunto seja amplamente revisado e estudado, visando a melhorias e adequações na execução dos testes.

Received on 09/19/19

Accepted on 10/24/19

INTRODUCTION

Evaluation is a vital activity for education planning and development. This activity allows following the process of acquisition and complexification of knowledge, assessing the evolution of learning that is relevant and significant for the development of students' skills and competences¹.

The Progress Test (PT) technique was initially developed in the early 1970s at the University of Missouri-Kansas City School of Medicine, in the United States, and at the University of Limburg, in the Netherlands².

Multiple-choice questions constitute one of the most widely used methods in tests designed to assess cognitive skills. The evaluation process aims to understand the adequacy between the students' performance and the year in which they are enrolled, in addition to fostering learning and supporting decisions that will have implications for their progress, contributing to the quality control of educational programs³.

The PT consists of an objective assessment, comprising 60 to 150 multiple-choice questions, aiming at promoting an assessment of the cognitive skills expected at the end of the course. This test is applied to all students on the same date, so that one can compare the results between the years and analyze the development of knowledge performance throughout the course⁴. The comparison of results is possible with the equalization of the tests⁶.

The maximum test time is 4 hours, and the minimum time for handing back the test is 1 hour. Depending on the institution, the presence of students may or may not be mandatory for the test to be carried out⁵.

Cognitive development is an important dimension in the training of

physicians, being a continuous process of acquisition and consolidation of a set of components necessary to master knowledge in one or more areas of performance⁶.

The tests designed to assess clinical competence in medical students need to be valid and reliable. It is acknowledged by the medical students' learning behavior that the assessment usually leads to learning. Therefore, if students are learning what is being assessed, it is vital that the assessment content reflect the learning objectives⁷.

The validity of an assessment is guaranteed when it is possible to demonstrate that the chosen method evaluates exactly what the student was intended to learn. It depends, then, on the adequacy of the method to the nature of the domain to be evaluated³.

Aiming to improve the quality of the tests, psychometric evaluation methods are analyzed, among them: the Classic Test Theory (CTT) and the Item Response Theory (IRT). The Classic Test Theory seeks to explain the total final result, considering the sum of the answers given at the test. This model aims to analyze two of the item parameters: the degree of difficulty and of discrimination. The analysis is based on the total score of right and wrong answers, without considering the students' latent skills and casual correct answers. Therefore, it evaluates the quality tests, assessing the correlation between the correct answers in a given item and the total number of correct answers in the test⁶.

In turn, the Item Response Theory (IRT) model is not interested in the total score of a test, but in the result obtained for each of the items that comprise the test. It aims to determine the probability and which factors affect the student's probability of giving a correct or wrong answer to an item of a certain test. Three parameters of the item are considered:

degree of difficulty, discriminative capacity and probability of casual correct answers. It also considers the individual's latent skills, seeking to analyze quality items⁶.

Embretson⁸ and Andrade⁹ emphasize that the IRT is the most recommended model to be applied in objective test evaluations, when compared to CTT, because it considers the students' latent skills and analyzes the parameters per item and not the test as a whole. IRT is also the method that best estimates the capacity of a certain item to have received a correct answer at random. In the progress test, this is a factor of great importance, since, as it is a test developed for the sixth year level, which involves several skills for solving the items, such as clinical reasoning, it is likely that beginner students will try to "guess" most of the items⁶.

The cumulative assessment method has been used as a tool to guide the students' study behavior, as it encourages students' self-study time more uniformly throughout the course¹⁰. Regardless of the philosophical stance of the evaluation, there is a common axis between the different conceptions of educational evaluation: they are associated with an interpretation process, assuming the judgment of value, quality and/or merit aiming at diagnosing and verifying the scope of the objectives proposed in the teaching-learning process¹¹. Despite the differences in the creation of the questions, the essence of the theories is to improve the teaching system and student learning.

OBJECTIVES

The study aimed at understanding the benefits of implementing the Progress Test in undergraduate medical school for the development of the learning, for the student, as well as for the teachers and the institution.

METHODOLOGY

Search strategy

A systematic and literary review was carried out about the Progress Test in medical schools in Brazil and worldwide. There was no language limitation for article inclusion in the study. Search methods were employed using health descriptors (DeCs Terminology) indexed to the database of the Latin American and Caribbean Center on Health Sciences Information (BIREME). The LILACS (Latin American Literature in Health Sciences), SCIELO (Scientific Electronic Library Online) and MEDLINE / PUBMED (US National Library of Medicine of National Institute of Health) databases were consulted.

All identified articles were read when the number was less than 20 results. Repeated articles were considered as those that had already been included in the study at a previous search.

Descriptors

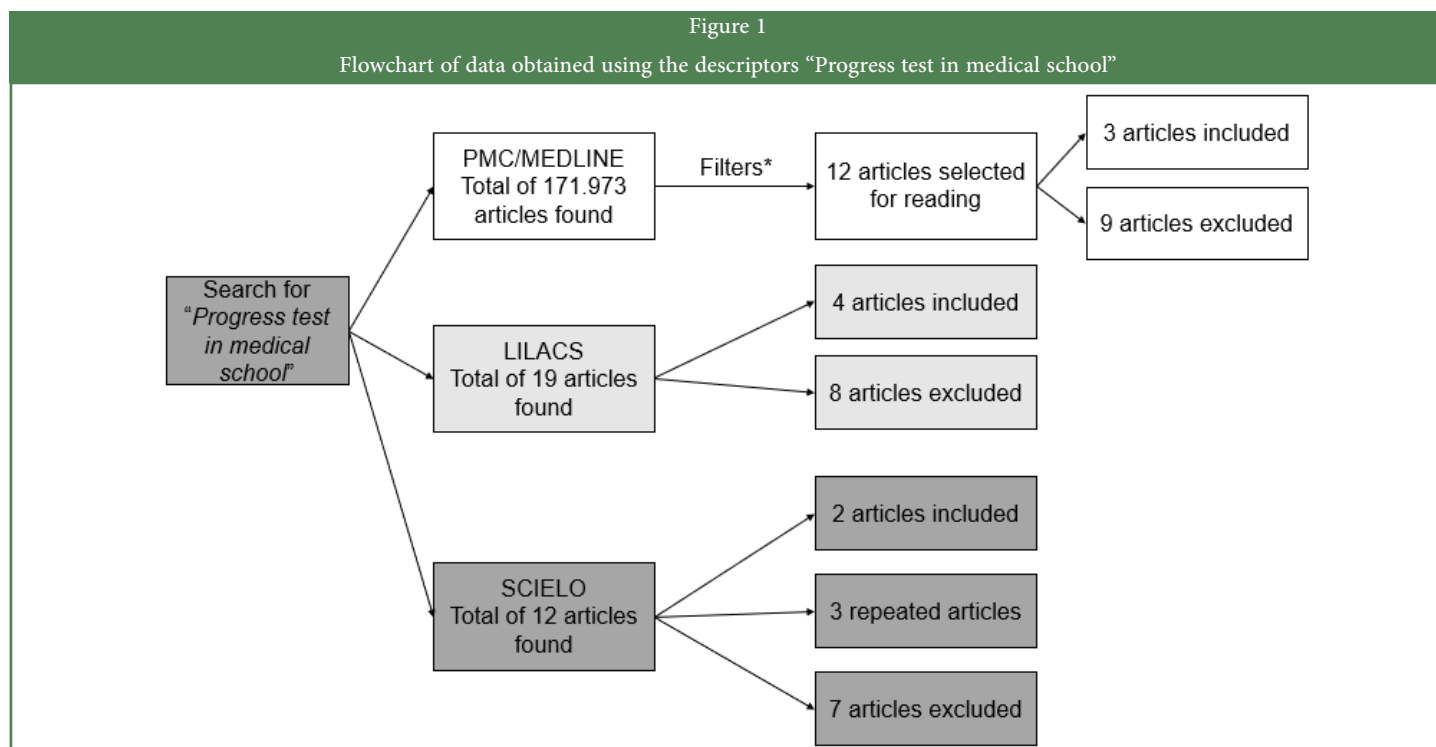
The descriptors used in the study were: "Progress Test in Medical Schools" and "Item Response Theory; Medicine". All searched terms were in the English language.

Search period

The search was carried out between July 2018 and April 2019, including articles that were published in the period from January 2002 to March 2019. The articles' date of publication was determined before starting the search.

Inclusion and exclusion criteria

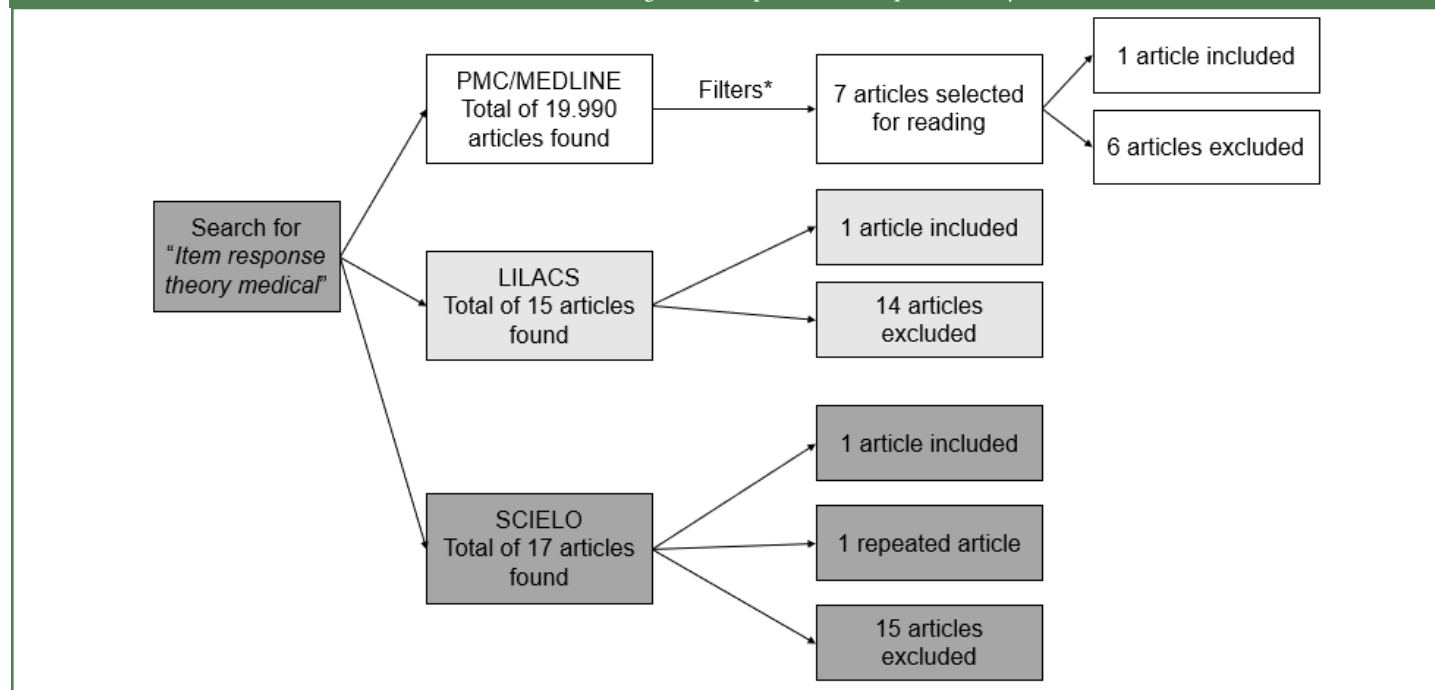
The inclusion criteria considered were (i) original articles; (ii) articles in which the topic was addressed without flight of ideas; (iii) articles that had been published within the pre-established period; (iv) articles



Source: the author. *Filters: Publication period from January 1, 2002 to March 31, 2019, with open access and from "MEDLINE journals"; then the "Search fields" filter was applied so that articles containing the descriptors in the body of the study were selected.

Figure 2

Flowchart of data obtained using the descriptors “Item response theory; medical”



Source: the author. * Filters: Publication period from January 1, 2002 to March 31, 2019, with open access and from “MEDLINE journals”; then the “Search fields” filter was applied so that articles containing the descriptors in the study title were selected.

with scientific contributions to the topic. The following were excluded: (i) studies that involved more than one school, and did not provide individualized prevalence estimates for each one of them; (ii) studies carried out at institutions of which data of interest to the research were not available from any online source (the institution’s website, Ministry of Education and Culture, or others); (iii) studies that did not address the proposed topic.

PubMed Central/MEDLINE

A total of 171,973 articles were found on the PMC platform (PubMed Central) for the descriptors “Progress test in medical school” with the publication period limitation between January 1, 2002 and March 31, 2019, with open access and from “MEDLINE journals”. The “Search fields” filter was applied in order to select articles containing the descriptors in the body of the study; 12 articles were found. Three articles were included. The other 9 articles were read and excluded; although they contained the topic “Progress Test” in the body of the study, they focused on other subjects that did not fit the proposed review.

The articles included were: “The use of progress testing” by Schuwirth and Van Der Vleuten (2012)¹²; “Development of a competency-based formative progress test with student-generated MCQs: Results from a multi-centre pilot study”, by Wagener et al. (2015)¹³; and “Progress testing in the medical curriculum: students approaches to learning and perceived stress” by Chen et al. (2015)¹⁴.

Still on the PMC platform, 19,990 articles were found for the descriptors “Item response theory; medical” with the same limitation of publication period, open access and from “MEDLINE journals”. After applying the same filters, 1,234 articles were found. The “Search fields”

filter was then used to select the articles that contained the descriptors in the title of the study, and 7 articles were identified. Among these, 1 article was included in the study. The other 6 articles were read and excluded, as they did not fit the review proposal.

The included article was: “Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education”, by Bhakta et al. (2005)⁷.

LILACS

The LILACS platform was used to search for the descriptors “Progress test, medical education”, and 19 articles were found for the publication period limited to January 1, 2002 to March 31, 2019. The articles were read, and of these, 4 articles were included in the study. The other 15 articles did not address the intended topic.

The articles included were: “*Avaliação cumulativa: melhora a aquisição e a retenção do conhecimento pelos estudantes?*” by Fernandes et al. (2018)¹⁰; “*Teste de Progresso: uma ferramenta avaliativa para a gestão acadêmica*” by Pinheiro et al. (2015)¹⁵; “*Avaliação do crescimento cognitivo do estudante de medicina: aplicação do teste de equalização no teste de progresso*” by Sakai, Ferreira Filho and Matsuo (2011)⁶; “*Teste de progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina*” by Sakai et al. (2008)².

The same platform was used to search for the keywords “Item response theory; medical” and 15 articles were found within the same publication period. The articles were read, and 1 article was selected for inclusion in the study; however, it had already been included in a previous search. The other 14 articles were not focused on the proposed topic.

SCIELO

The SCIELO platform was used to search for the keywords “Progress test, medical education”, and 12 articles were found within the publication period limited to January 1, 2002 to March 31, 2019. The articles were read and 2 of them were included in the study. In addition to these, there were 3 repeated articles. The other 7 articles did not address the studied topic.

The articles included were: “O Teste de Progresso como indicador para melhorias em curso de graduação em medicina” by Rosa et al. (2017)¹⁶; “Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a elaboração de testes de múltipla escolha e a composição de exames”, by Bollela et al. (2018)³.

The same platform was used to search for the keywords “Item response theory; medical” and 17 articles were found. All articles were read, and two were selected according to the inclusion criteria; however, one of them had already been included in the study (repeated). The other 15 articles were excluded because they did not fit the proposed topic.

The included article was: “Avaliação critério-referenciada em medicina e enfermagem: diferentes concepções de docentes e estudantes de uma escola pública de saúde de Brasília, Brasil” by Miranda Júnior et al. (2018)⁴.

RESULTS

Eleven scientific papers on the topic were included (Table 1).

The studies included helped to understand the concept of Progress

Test and of cognitive development in general, mainly related to the construction of medical knowledge.

Schuwirth and Van der Vleuten (2012) carried out a literature review about the use of PT. It was demonstrated that the longitudinal assessment approach has a positive effect on the students’ learning behavior, discouraging pre-test learning, and consequently helping in the long-term cumulative learning, the long-lasting learning. It is assumed that students experience a lower level of stress when performing the PT in comparison to the traditional tests, since a single bad result cannot undo a series of good results. The authors also emphasize that the longitudinal method increases the reliability of the PT¹².

According to Wagener et al (2015), the PT provides students with feedback on their level of proficiency throughout their studies. The PT was applied to students from the 1st to the 6th years of the Medicine course in seven medical schools in Germany, totaling 463 participating students. Of these 61.3% were females, 35% males and 3.7% had no gender indication. The mean age of the participants was 24.56 years. The study observed that the number of correct answers constantly increased with the advance in the level of the academic semester in which the student was enrolled¹³.

Chen et al (2015) explored the effect of PT compared to traditional tests on the stress perceived by the students. The study was carried out in two stages, evaluating the change in relation to the elapsed time. There were a total of 864 participating students. Throughout the two assessed moments, the PT may have reduced the stress for students, as those who

Table 1

Articles included in the study: authors, country where the study was carried out, year of publication and study type.

Title	Authors/Country	Year of publication	Study type
The use of progress testing	Schuwirth and Van Der Vleuten / The Netherlands	2012	Literature review
Development of a competency-based formative progress test with student-generated MCQs: Results from a multi-centre pilot study	Wagener et al / Germany	2015	Cohort
Progress testing in the medical curriculum: students approaches to learning and perceived stress	Chen et al / New Zealand	2015	Multicenter pilot study
Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education	Bhakta et al / United Kingdom	2005	Cohort
Cumulative assessment: does it improve students’ knowledge acquisition and retention?	Fernandes et al / The Netherlands	2018	Cohort
Teste de Progresso: uma ferramenta avaliativa para a gestão acadêmica	Pinheiro et al / Brazil	2015	Cohort cross-sectional
Avaliação do crescimento cognitivo do estudante de medicina: aplicação do teste de equalização no teste de progresso	Sakai et al / Brazil	2011	Action Research Test
Teste de progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina	Sakai et al / Brazil	2008	Cross-sectional observational-aggregate
O Teste de Progresso como indicador para melhorias em curso de graduação em medicina	Rosa et al / Brazil	2017	Cross-sectional cohort
Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a elaboração de testes de múltipla escolha e a composição de exames	Bollela et al / Brazil	2018	Literature review
Avaliação critério-referenciada em medicina e enfermagem: diferentes concepções de docentes e estudantes de uma escola pública de saúde de Brasília, Brasil	Miranda Júnior et al / Brazil	2018	Quali-quantitative, descriptive study with cross-sectional design

Source: the author.

participated in the traditional group experienced significant increases in stress when they took the traditional year-end exams, while stress levels did not significantly increase for the students in the PT group¹⁴.

Bhakta et al (2005) analyzed the results of the Extended Matching Questions Examination carried out by students of the 4th year of medical school (between 2001 to 2002). Rasch analysis was used to assess whether the set of questions used in the examination corresponded to a one-dimension scale. The degree of difficulty of medical and surgical intra and inter-specialty questions was observed, in addition the pattern of answers in the individual questions to evaluate the impact of the distractor options. The study allowed the understanding of how students use the provided options and information, together with specialized knowledge, to understand the questions and decide on the correct answer. This information is essential for the editors to improve the quality of the questions to be created⁷.

Fernandes et al. (2018) studied the cumulative assessment as a tool to guide the students' study behavior, comparing the increase in knowledge between students who participated with cumulative assessment and those who remained in the traditional method of assessment at the end of the teaching cycles. Data from the first four Dutch inter-university PTs prior to the experiment were used. A total of 62 students participated; among these, 37 underwent the traditional assessment and 25, the cumulative assessment. It was demonstrated that there was a significant increase in students' knowledge in the four PTs throughout the course. It was not possible to show any difference between the two groups, indicating that the two assessment methods are similar in relation to the increase in students' knowledge¹⁰.

Pinheiro et al (2015) investigated the potential of PT in aiding academic management. The students' performance was analyzed, from the first to the sixth years of medical school, in the PT applied in 2008 and reapplied without modifications in 2011. The results showed that at the two moments of test application, the students acquired cognitive knowledge during the medical course. However, the cumulative learning behavior showed differences between the two moments of test application. In 2008, the progression occurred every two years of the course and in 2011 this progression started only after the third year of the course. This study was crucial for managers to understand the need for greater considerations on the teaching offered and the quality of the questions created for the assessment¹⁵.

An assessment of the medical students' cognitive development was carried out by applying the Equalization Test to the Progress Test at a Brazilian state university by Sakai et al (2011). All results of students from the first to the sixth academic years in the period from 2004 to 2007 were analyzed. At first, they tried to explain the meaning of the responses given by students to a series of items (questions). Two psychometric models were used: CTT and IRT. For the analysis with the CTT, one good-quality item was considered to have a biserial correlation (between the correct answer for a certain item and the total number of correct answers in the test) >0.40. For the IRT, three parameters were used for multiple-choice items: (a) the item's discrimination capacity; (b) the degree of difficulty of the item; (c) random correct answer⁶.

Still regarding the study by Sakai et al (2011), it was not possible to proceed further into the IRT application stage, since the number of items with a biserial correlation >0.40 was below the recommended level,

in addition to the insufficient number of necessary respondents per group. However, it was possible to define the anchor-questions that would be used in the next moment. The mean number of correct answers given by the students, in all PT, varied from a minimum of 33.5% in the first year, to a maximum of 66.4% in the sixth year, in the results without equalization. As for the mean scores in the PT with equalization, they ranged from 31.0% to 73%, for the first and sixth years, respectively. It was suggested that medical schools establish partnerships to constitute a larger base of questions and experiences with pre-established criteria, increasing the number of items with a high degree of discrimination and respondents, which would allow using the IRT for the equalization of tests⁶.

In 2008, Sakai et al. carried out an evaluation of the medical course and the use of PT during 10 years of experience at a state medical school in southern Brazil. The PTMed was used as an instrument to evaluate the course and as a result of this period there was an increase in the participation of students to take the test; the students' cognitive performance increased from one year to another in each test; the mean number of correct answers in the areas of Clinical Medicine was lower in the fifth year, in the eighth and ninth PT; in Public Health, there was a high percentage of correct answers in the first two years. Such results reflect the curricular structure, as well as the strengths and weaknesses of the course. This study was a good indicator of the self-assessment process of the course, but it was suggested that further investigation was still necessary regarding the studies of result analysis techniques to estimate the students' cognitive development².

Rosa et al. (2017) used the Progress Test as an indicator for improvements in the undergraduate medical course. Three cross-sectional institutional studies were carried out during the three years of test application in the medical course at a university in the southern Brazil. Undergraduate medical students participated and had taken the progress test in 2011, 2012 and 2013. The statistical analysis was carried out with a 95% confidence level. The mean adherence over the three years ranged from 91.8% to 100%. It was concluded that the PT is an excellent indicator for managers, as it can be used to develop interventions to improve course quality. After the first test was applied, changes were made to the medical course at the university. Subsequent tests demonstrated the effectiveness of these changes¹⁶.

Bollela et al. (2018) used a summative assessment of cognitive skills to understand more about the use of good practices for the development of multiple-choice tests (questions) and the structure of exams. The author mentions that for the assessment of skills related to the cognitive domain, three reference levels are used to create the questions: (a) basic, which requires factual knowledge, basically involving the memorization of facts and their recovery; (b) intermediate, which involves not only the acquired knowledge but also the skills to understand, interpret and apply that knowledge to solve simpler problems; (c) advanced, which involves the levels of analysis, synthesis and evaluation to propose solutions to more complex problems, such as those that are usually presented to health professionals. The study concluded that multiple-choice questions (MCQ) with only one correct alternative have several advantages, but for their adequate applicability, it is necessary to pay attention to good practices when creating questions and building exams, involving recommendations related to the test content and format, as well as the composition of the evaluation matrix or specification table³.

Miranda Junior et al. (2018) studied different conceptions of teachers and students from a public health school in Brasília about the criterion-referenced evaluation in Medicine and Nursing. This type of assessment analyzes academic performance without comparing one student to another, differing from the standardized assessment that aims to determine the best and worst performances. A sample of 413 participants was obtained, totaling 54 teachers and 344 students. A weakness in the concept of criterion-referenced evaluation was observed in the entire academic body and dissatisfaction regarding the evaluation criteria adopted at school (323, or 79.6%). Among the indicated reasons, the subjectivity in the exam correction used and difficulties when applying for residency positions stand out. Other authors report that the criterion-referenced method tends to inflate students' references, although it evaluates them in a more comprehensive way. For this reason, they recommend the adoption of the criterion-referenced and standardized systems to, respectively, better measure student performance and compare them with each other. The author concludes that taking into account all this, it is necessary to adopt both models (standardized and criterion-referenced), with the application of performance standards. The assessment instruments also need more objective criteria for a more uniform and just test correction⁴.

DISCUSSION

The Progress Test aims to assess students' cognitive performance during the medical course, as well as being an instrument that helps in understanding the successes and failures of the course itself. A single test is developed, and it is applied to all semesters / years of the course at the same time². When applied annually, it allows verifying student evolution throughout the course.

It is a type of longitudinal assessment of the students' cognitive development during the course. The tests applied by the different medical schools have differences among them, such as: standardized or criterion-referenced; questions can be of the true / false or multiple-choice type¹⁷.

In the medical education scenario in Brazil, the Progress Test in Medicine (PTMed) has been applied in medical schools, either alone or in partnership, since the late 1990s⁶.

Considering the interest of medical schools in the development of educational process, partnerships were created to share information and learning, making it possible to improve the application and usefulness of the test in medical education. With this objective in mind, the South II Interinstitutional Pedagogical Support Center (*Núcleo de Apoio Pedagógico Interinstitucional Sul II*, NAPISUL II) was established in 2010 through the initiative of eight medical schools with support from the Brazilian Medical Education Association (*Associação Brasileira de Educação Médica*, ABEM), with the objective of creating, implementing and analyzing the Progress Test¹⁶.

Moreover, in 2004 the Ministry of Education introduced a new model for the National Student Performance Exam (*Exame Nacional de Desempenho dos Estudantes*, ENADE), of which objectives are similar to those of the Progress Test but applied only to students attending the first and last years of the courses¹⁸. The Progress Test is an institutional assessment, which does not aim at creating a ranking among medical schools; therefore, the test dates are defined by each institution, center or consortium¹⁹.

The PTMed questions are prepared by a multidisciplinary team, consisting of teachers from the institutions themselves, and have different

degrees of difficulty. Each question has four to five possible alternatives as answers. Moreover, they are divided into seven areas of knowledge: General Practice, Pediatrics, Gynecology and Obstetrics, Surgery, Public Health, Medical Ethics and Basic Sciences. The questions can be contextualized in clinical cases, so that the statements require not only the students' memorization, but also their logical reasoning¹⁹.

The test results build academic performance curves, which allows identifying weaknesses and strengths of students in the several areas of knowledge related to the course².

The evaluation methods can be used as a tool for improving teaching. Formative assessment can be defined as continuous evaluation, of a non-probative nature, aimed at improving learning. The evaluation process acts as a regulator, showing failures and possible solutions to obstacles presented by the students, which can lead to improvements in didactic tools¹⁹.

The IRT became known mainly from 1968, with the work of Lord and Novick entitled "Statistical Theories of Mental Tests Scores". Since then, several applications have been explored for the theory: creation of item banks, computerized adaptive assessment, equalization of tests, cognitive change assessment, among others. The IRT has the item as its unit of analysis, formalizing the relationship between the probability of getting right answer for the item and the latent capacity required to resolve it. The greater the capacity of a subject (latent trait), the greater the probability that this subject will get the right answer to a certain item that measures this construct²⁰.

The item response theory (IRT) has been considered a milestone for modern psychometry, showing advantages over the Classic Test Theory (CTT), considering the virtual invariance of the item parameters in relation to the sample, a more accurate and interpretable estimation of individuals' skill levels and more efficient test equalization procedures²⁰.

The Progress Test can also be an auxiliary instrument in training students for the selection processes in which they will participate during their professional life, such as applying for medical residency positions².

According to the degree of difficulty, the question is considered easy when there is a rate of correct answers $\geq 80\%$, intermediate when the rate is between 20% and 80% and difficult when there is a rate of correct answers $\leq 20\%$ ⁶.

The trend for medical schools to establish partnerships to prepare and apply the tests indicates test quality improvement, as this allows the exchange of experiences between them, the creation of working groups with the associated specialties and the introduction of an item bank with tested and evaluated questions. Additionally, the number of items with a high degree of discrimination and respondents would increase, which would allow the use of IRT for the equalization of tests⁶.

The assessment of student performance depends fundamentally on the quality of the test items. Therefore, CTT and IRT are statistical methods that can contribute to this purpose²¹.

The Progress Test is not an exclusive instrument for assessing student performance, but it is also important as an assessment tool for use in academic management and, for this purpose, it is essential that institutions take an active role in the creation and data analysis of the assessment¹⁵.

Another assumed advantage is that the collection of longitudinal data is more predictive of competence / future performance than single measurements.

Among the advantages of applying the test, it can be observed that

the predictive validity is increased as reliability increases, and this is due to the longitudinal combination of results. The collection of longitudinal data is more predictive of future competence / performance when compared to single measurements¹².

The possibility of determining the characteristics of difficulty, discrimination and random correct answers encourages new forms of evaluation; thus, Computerized Adaptive Progress Tests are being developed^{23,24}.

The PTMed allows the course to review its pedagogical project and contents based on general analyses and per area. With its routine use, data analysis can provide courses and students with an external cognitive assessment, as there is no endogeny resulting from questions always made by the same teachers²⁵.

Educational changes begin in small steps: the union of medical schools can assist, implement and optimize an economic assessment instrument capable of promoting improvements in learning. In addition to being strategic for learning and management, the PT is a low-cost instrument²².

Moreover, the correlation between the students' progress test scores and their performance in a residency selection process was assessed, which demonstrated that indeed, there is a correlation between the students' performance in PTMed and in the residency test. These data can be explained by the fact that both instruments measure cognitive skills and knowledge²⁶.

FINAL CONSIDERATIONS

This study main contribution was the discussion about the concept of learning during the time spent by students in medical school.

The interest in educational progress led to the development of studies about learning. The Progress Test in Medicine (TP) emerged as a way to measure the students' development in relation to the elapsed time attending the course. Additionally, the test allows identifying probable deficiencies in the active subjects of the course, aiming to improve and intervene in teaching methods. The students benefit by quantifying their learning progress over time and assisting them in preparing for the medical residency selection tests. As for the institution and its managers, it is possible to determine points of deficiency and course efficiency, allowing improvements in the teaching methods.

More quantitative studies are suggested for better statistical analysis of the quality in the creation of cumulative tests and determining their positive and/or negative impact on the construction of learning.

REFERENCES

1. Ferreira ARL, Weyh CB. Tecnologia da informação no processo da avaliação como estratégia para qualificar a prática docente. In: Congresso Internacional de Educação e Tecnologias; 2018; São Carlos, Brasil. São Carlos: UFSCar; 2018. p. 1-7.
2. Sakai MH, Ferreira Filho, OF, Almeida MJ, Mashima DA, Marchese MC. Teste de Progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina. *Rev. bras. educ. med.* 2008;32(2):254-63.
3. Bollela VR, Borges MC, Troncon LEA. Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a

- elaboração de testes de múltipla escolha e a composição de exames. *Rev. bras. educ. med.* 2018;42(4):74-85.
4. Miranda Junior UJP, Tatsch JFS, Braga CS, Spengler Junior SL, Novaes MRCG. Avaliação critério-referenciada em Medicina e Enfermagem: diferentes concepções de docentes e estudantes de uma escola pública de saúde de Brasília, Brasil. *Rev. bras. educ. med.* 2018; 42(3):67-77.
5. Associação Brasileira de Educação Médica. Avaliação do estudante: contribuição do teste de progresso. ABEM, 2017. Disponível em: <https://abem-educmed.org.br/acoef/avaliacao-do-estudante-contribuicao-do-teste-de-progresso/>. Acesso em: 14 de maio de 2019.
6. Sakai MH, Ferreira Filho, OF, Almeida MJ, Matsuo T. Avaliação do crescimento cognitivo do estudante de Medicina: aplicação do teste de equalização no Teste de Progresso. *Rev. bras. educ. med.* 2011,35(4):493-501.
7. Bhakta B, Tennant A, Horton M, Lawton G, Andrich D. Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education. *BMC med. educ* 2005;5(9):1-13.
8. Embretson SE. The new rules of measurement. *Psychol. Assess.* 1996;8(4):341-9.
9. Andrade DF. Comparando desempenho de grupos de alunos por intermédio da teoria de resposta ao item. *Est. aval. educ.* 2001;23:31-69.
10. Fernandes DC, Nagtegaal M, Noordzij G, Tio RA. Cumulative assessment: does it improve students knowledge acquisition and retention? *Scientia Medical Education in Health Sciences* 2018;28(4):ID31880.
11. Rodrigues M. Proposta de análise de itens das provas do Saeb sob a perspectiva pedagógica e a psicométrica. *Est. aval. educ.* 2006;34:43-78.
12. Schuwirth LWT, Van Der Vleuten CPM. The use of progress testing. *Perspect Med Educ* 2012; 1:24-30.
13. Wagener S, Moltner A, Timbil S, Gornostayeva M, Schultz JH, Brustle P et al. Development of a competency-based formative progress test with student-generated MCQs: results from a multi-centre pilot study. *GMS Z Med Ausbild* 2015;32(4):Doc46.
14. Chen Y, Henning M, Yelder J, Jones R, Wearn A, Weller J. Progress testing in the medical curriculum: students approaches to learning and perceived stress. *BMC Med Educ* 2015; 15(147):1-8.
15. Pinheiro OL, Spadella MA, Moreira HM, Ribeiro ZMT, Guimarães APC, Almeida Filho OM et al. Teste de Progresso: uma ferramenta avaliativa para a gestão acadêmica. *Rev. bras. educ. med.* 2015;39(1):68-78.
16. Rosa MI, Isoppo CC, Cattaneo HD, Madeira K, Adami F, Ferreira Filho OF. O Teste de Progresso como indicador para melhorias em curso de graduação em Medicina. *Rev. bras. educ. med.* 2017;41(1):58-68.
17. Verhoeven BH, Verwijnen GM, Scherpbier AJ, van der Vleuten CP. Growth of medical knowledge. *Med Educ* 2002;36(8):711-7.
18. Limana A, Brito MRF de. O modelo de avaliação dinâmica e o desenvolvimento de competências: algumas considerações a respeito do Enade. *Avaliação* 2005;10(2):9-32.
19. Baldim TL, Vicente CAO, Arcuri MB. O teste de progresso sob a visão do discente. *Revista da Faculdade de Medicina de Teresópolis* 2018;2(1):41-54.
20. Nunes CHSS, Primi R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item. *Aval. psicol.* 2005;4(2):141-153.

21. Henning E, Ramos MS, Konrath AC. Análise de itens de uma prova de raciocínio probabilístico. In: Actas del VII CIBEM; 2013; Montevideo, Uruguay. Colegio Seminario; 2013. p. 2025-2032.
22. Vleuten CVD, Freeman A, Collares CF. Profess test utopia. *Perspect Med Educ* 2018;7:136-8.
23. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. *Med Educ* 2010;44(1):109-17.
24. Collares CF, Cecilio-Fernandes D. When I say...computerised adaptative testing. *Med Educ* 2019;53(2):115-6.
25. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner MLMB, Maffei CML. Consortia of cross-institutional progress testing for all medical schools in Brazil. *Rev. bras. educ. med.* 2019; 43(4)151-6.
26. Hamamoto Filho PT, Lourenção PLTA, Valle AP do, Abbade JF, Bicudo AM. A correlação entre a pontuação dos testes de progresso dos alunos e seu desempenho em um processo de seleção de residência. *Medical Science Educator* 2019;29(1):1-5.

AUTHORS' CONTRIBUTIONS

Ademir Garcia Reberti participated in the research, selection of texts, writing and corrections about the manuscript.

Nayme Hechem Monfredini participated in the research, selection of

texts, writing and corrections about the manuscript.

Olavo Franco Ferreira Filho participated in the research, reading, indicating which articles were most relevant for the making of this manuscript.

Dalton Francisco de Andrade participated in the research, reading, indicating which articles were most pertinent for the making of this manuscript, especially with regard to statistics.

Carlos Eduardo Andrade Pinheiro participated in the research, reading, indicating which articles were most pertinent for the making of this manuscript, especially with regard to the Progress Test at the international level.

Jean Carl Silva participated in the research, selection of texts, writing and corrections about the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ADDRESS FOR CORRESPONDENCE

Ademir Garcia Reberti

Rua Henrique Meyer, 280, Centro, Sala 311, Joinville, Santa Catarina, Brazil. CEP: 89201-405.

E-mail: agrebert@terra.com.br



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.