

## Original Article

## 10 years of experience in adopting, implementing and evaluating progress testing for Saudi medical students



Ahmad S. Alamro, PhD<sup>a</sup>, Abdullah A. Alghasham, PhD<sup>b</sup>, Hani A. Al-Shobaili, PhD<sup>c</sup>, Homaidan T. Alhomaidan, MD<sup>d</sup>, Tarek A. Salem, PhD<sup>e</sup>, Majed M. Wadi, MSc<sup>a,\*</sup> and Mohamed N. Saleh, PhD<sup>f</sup>

<sup>a</sup> Medical Education Department, College of Medicine, Qassim University, KSA

<sup>b</sup> Pharmacology Department, College of Medicine, Qassim University, KSA

<sup>c</sup> Dermatology Department, College of Medicine, Qassim University, KSA

<sup>d</sup> Family Medicine Department, College of Medicine, Qassim University, KSA

<sup>e</sup> Pathology Department, College of Medicine, Qassim University, KSA

<sup>f</sup> Anatomy Department, College of Medicine, Qassim University, KSA

Received 25 February 2022; revised 5 July 2022; accepted 12 July 2022; Available online 10 August 2022

### المخلص

**الأهداف:** اختبار التقدم (أو ما يعرف بالاختبار التحصيلي) هو اختبار شامل مصمم لمطابقة اكتساب المعرفة اللازمة عند التخرج. يراقب هذا الاختبار التقدم خلال الفترة الكاملة لبرنامج البكالوريوس. بدأت كلية الطب بالقصيم بتصميم وتبني هذا الاختبار الذي يشارك فيه كليات الطب على مستوى في المملكة العربية السعودية. هدفت الدراسة الحالية إلى تحديد ما إذا كان يمكن استخدام هذا الاختبار لتقييم تقدم طلاب الطب في كليات الطب السعودية المختلفة بمقاربات تعليمية مختلفة، وكذلك ما إذا كان اعتماد طريقة الاختبار هذه ستقبله الكليات المختلفة.

**الأساليب:** إلى جانب إنشاء اللجنة الخاصة بالاختبار، تم وضع مخطط شامل لأخذ 200 سؤال من أسئلة الاختبار من من تخصصات مختلفة. الاختبار عبارة عن اختبار كتابي في أربع ساعات. جميع عناصر الأسئلة فيه عبارة عن أسئلة ذات خيارات متعددة مبنية على سيناريو.

**النتائج:** تم إجراء 13 جولة من اختبار التقدم. ارتفع عدد الكليات المشاركة من ثلاث كليات (285 طالباً) في الاختبار الأول (مايو 2012) إلى أكثر من 20 كلية (أكثر من 6000 طالب) من الجولة التاسعة (فبراير 2017). تراوح متوسط درجات طلاب السنة الأولى من 3.0% إلى 7.9%، بينما تراوحت درجات طلاب السنة الخامسة من 34.0% إلى 43.0%.

**الاستنتاج:** يعتبر إجراء هذا الاختبار المصمم بدقة لتقييم التحصيل المعرفي عند التخرج من الطب تجربة مثمرة؛ لقد ساعد في تقديم ملاحظات بناءة للمتقدمين للاختبار وأصحاب المصلحة الآخرين على المستوى الوطني.

**الكلمات المفتاحية:** اختبار التقدم؛ التقييم؛ التقييم التكويني؛ MCQs؛ طلاب الطب؛ الكفاءة الطبية

### Abstract

**Objectives:** The progress test (PT) is a comprehensive examination that is designed to match the knowledge acquisition necessary at graduation and monitors progress during the entire period of an undergraduate program. Qassim College of Medicine (QCM) began using the multi-institutional PT in the Kingdom of Saudi Arabia (KSA). This study aimed to determine if the PT can be utilized to assess the progress of medical students at different Saudi medical colleges with different educational approaches, as well as whether this testing modality could be accepted by other colleges.

**Methods:** Beside the establishment of a PT committee, comprehensive blueprinting was crafted to sample 200 A-type multiple choice questions (MCQs) from different disciplines. The PT is a paper-and-pencil model and is answered in a 4-h period. All PT items followed a uniform design.

**Results:** In total, 13 rounds of the progress test have been conducted. The number of participating colleges

\* Corresponding address: Majed Mohammed Wadi, Medical Education Department, College of Medicine, Qassim University, KSA.

E-mail: [m.wadi@qu.edu.sa](mailto:m.wadi@qu.edu.sa) (M.M. Wadi)

Peer review under responsibility of Taibah University.



increased from three (with 285 students) in the first test (May 2012) to more than 20 (with >6000 students) in the ninth round (February 2017). The average % scores for first-year students ranged from 3.0% to 7.9% while the average scores for fifth-year students ranged from 34.0% to 43.0%.

**Conclusion:** The conduction of this meticulously crafted test to evaluate knowledge achievement at medical graduation is a fruitful tool and helps to provide constructive feedback for test-takers and other stakeholders relating to their relative positions among other fellows at the national level.

**Keywords:** Assessment; Formative assessment; MCQs; Medical competency; Medical students; Progress test

© 2022 The Authors.

Production and hosting by Elsevier Ltd on behalf of Taibah University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

At the start of the 21st century, the establishment of medical colleges in the Kingdom of Saudi Arabia (KSA) rapidly expanded, aiming principally to increase the number of Saudi doctors as a part of the Saudization of healthcare services via the replacement of expatriates by natives in the health care workforce. Through the first decade of the century, the number of medical colleges markedly increased from 5 to 41, including both governmental and private colleges.<sup>1</sup> These newly established colleges adopted different educational approaches; some colleges apply an international college program as such or after adjustment to the Saudi context.<sup>2</sup> With this rapid expansion, queries were raised about the effectiveness of such approaches and their fitness to build up the needed competencies of graduates.<sup>3,4</sup>

In 2007, the “Saudi Future Doctors” project tried to determine the minimum competencies of Saudi medical graduates<sup>5</sup>; this project met a preliminary consensus that motivated the Committee of Deans of Saudi Medical Colleges to support and develop it into the SaudiMEDs framework that was published in 2011.<sup>3</sup> In parallel, as the Qassim College of Medicine (QCM) was in the initial stages of applying progress testing for its students, questions were raised about this longitudinal assessment modality and its utility in testing the progress of medical students of various colleges with different educational approaches. This proved to be true in many countries.<sup>6</sup>

A progress test (PT) is a comprehensive examination that samples the complete domain of knowledge acquisition necessary for a medical graduate to master. This test was introduced in the late seventies in many medical colleges adopting the problem-based learning (PBL) strategy in the Netherlands, Canada, and Germany.<sup>7–9</sup> The main rationale behind introducing the PT is the development of an assessment tool that measures knowledge achievement and

monitors its progress during the entire duration of a medical program.<sup>10</sup>

The main criteria of the progress test include independence of the curriculum stage, as multiple classes of students sit for the same test at the same time; repetition, either annually, or twice or more per year; and testing the use of knowledge rather than simple memorization, as there is no intensive pre-test revision.<sup>7,10,11</sup>

Previous studies have reported that the PT has helped medical colleges to monitor the progress of student learning, both individually and as groups, towards achieving the program learning outcomes (PLOs). Feedback from the PT gives students insight relating to their strengths and areas that need further work; it also informs each student about their own progress and position among peers in their college, as well as all participants from other universities. Furthermore, the PT can be a useful tool for benchmarking.<sup>10</sup> The association between future competence and/or the performance of graduates with proper monitoring, formative assessment, and feedback during undergraduate studies has already been reported.<sup>12</sup>

QCM was established in 2001 as the first medical college in KSA to adopt an integrated medical curriculum with PBL as its average strategy. The college vision is “a leading national and internationally recognized college in innovative medical education supporting the development of a healthy community”. On this basis, the college, since its commencement, has aimed to provide quality medical education for its own students<sup>13</sup> as well as contributing to the enhancement of medical education across KSA. In 2010 and 2011, the college ran two rounds of pilot progress tests for its own students. In May 2012, it began using the multi-institutional PT with the participation of three colleges. In November 2012, 12 colleges participated.<sup>6</sup> Up to March 2017, nine rounds were conducted with the participation of 25,270 students from 27 colleges.

The tenth round of the test (March 2018) had a fundamental change in the blueprint to match the SaudiMEDs framework themes and PLOs,<sup>14,15</sup> in which the expected competencies of Saudi medical graduates are determined according to six themes. This blueprint also outlines the required knowledge and skills to build up these competencies throughout the learning journey. Furthermore, the PT blueprint considered the expected activities of physicians and dimensions of care adopted by the Saudi Commission For Health Specialties (SCFHS) in the Saudi Medical Licensing Examination.<sup>16</sup>

The principal aim of the current study was to determine whether progress testing can be utilized to assess the progress of medical students of various Saudi medical colleges with different educational approaches, as well as whether the adoption of this testing modality would be accepted by various colleges.

## Materials and Methods

Implementation of the PT was accomplished through the following eight phases.

### *Phase 1: Progress test committee*

The QCM established the Progress Test Committee, headed by the Vice Dean for Educational Affairs and

including scientific and logistic subcommittees. The first committee is responsible for setting the test blueprint, the quality criteria for PT items and ensuring that items are matched to the blueprint. Communicating with other colleges in KSA, determining the test date to avoid national or academic breaks and suit most colleges, and sending and receiving test papers and results are among the duties of the logistic subcommittee. In addition, it the logistic subcommittee conducts appropriate orientation for faculty and students about the test, including rationale, nature and expected advantages. This includes interactive lectures, discussion forums, and two rounds of pilot PTs so that students and faculty gain a realistic experience of the process.<sup>17</sup>

#### *Phase 2: Benefit gained from the experience of others*

Cooperation with Maastricht University (MU) was the next step. This played an essential part in preparation and orientation plans. Through an official contract, experts from MU supported the Medical Education Department of QCM to conduct many sessions to promote the test, prepare and select items top test, and utilize data arising from test results.

#### *Phase 3: The determination of test features*

The test was constructed with 200 A-type MCQs to be completed within four continuous hours and facilitated as a paper-and-pencil test once or twice per year according to the pre-determined dates approved by the participant colleges. Nevertheless, building up a comprehensive multi-dimensional blueprint was the real foundation of the PT.

#### *Phase 4: Constructing the test blueprint*

The multifaceted blueprint of the PT considered many aspects to ensure complete coverage and appropriate sampling of the cognitive domain of the medical program. The blueprint targeted three main themes: body systems (e.g., respiratory, digestive, cardiovascular, nervous, locomotor), medical disciplines (e.g., anatomy, physiology, surgery, family medicine, psychiatry), and processes that graduates are expected to understand (e.g., mechanism of disease, diagnosis, management, therapeutics, health maintenance, psychosocial aspects). The distribution of items per theme was determined by consensus of the PT scientific committee, assessment experts, the Medical Education Department, and representatives of medical disciplines.

the “do not know” option was to promote professionalism by encouraging students to admit lack of information when relevant.<sup>18,19</sup>

The PT committee at QCM set rigorous criteria for PT items; the main issues addressed were common health problems in the Saudi context and identifying crucial scenarios for early intervention. Other criteria for writing credible A-type MCQs were included, and inclusion of the needed data in the vignette, clarity of phrasing with no redundancy and categorization of items to each of the themes of the blueprint were highlighted. The instructions for designing PT items were sent to all discipline departments and hands-on training sessions for writing credible A-type MCQs were continuously conducted. Disciplines were then asked to design twice as many items as needed according to the test blueprint. The prepared items were subject to a meticulous review process: after first approval at the departmental level, items were scrutinized through meetings of the scientific committee, including area and non-area experts and medical educationists. As per the preset criteria, items were either approved, rejected, or modified. Finally, those matching the test blueprint were selected.

#### *Phase 6: Test logistics*

The test date was preliminarily planned accounting to the educational calendar of the Ministry of Education and was fixed or modified based on communication with other colleges of medicine in KSA. All medical colleges in KSA were invited to participate in the test and fill out a specific form to document the expected number of their students who would take the test. Accordingly, QCM sent specific optical mark reader papers (OMRs) to each college according to the number of students in advance. Additionally, they were sent clear instructions and guidance on test issues to be given to students.

#### *Phase 7: Test results*

##### *Calculation of score*

To be a safe medical practitioner, it is important not to rush diagnosis or to suggest management options without being reasonably sure of the appropriateness. To minimize unjustified guessing, results are recalculated considering the number of both correct and incorrect answers. Thus, examinees were penalized for wrong answers, the score (%) in the whole test and each particular component (category) was calculated as follows:

---


$$\text{Score (\%)} = \frac{\text{Number of correct answers} - 0.25 \times \text{Number of wrong answers}}{200} \times 100$$


---

#### *Phase 5: Creating test items*

Test items were prepared *de novo* by QCM. All items followed a uniform design: a statement with a vignette, a lead-in, and five options (the correct choice, three distracters, and “do not know”). In addition, to reduce guessing, which would reduce the reliability of the test, the main rationale of

The PT items targeted the PLOs that students were not expected to know in the early years of their program to achieve that level of knowledge. Thus, “do not know” was included as an option for every item, but without penalty for choosing it or for not responding. This was to ensure students acknowledged if they did not know and to promote their professionalism.<sup>19,20</sup> Additionally,

the difficulty index for each test item was separately calculated for each program level to show with advancement in the program.

#### Reporting of results

*Students.* Results were uploaded using the college learning management system along with a video showing how students can access their results. Their usernames were extracted from their student data, and every student had their own password that they had already decided upon and was included in the specific region of the OMR sheet. Technical support for those encountering difficulties accessing their results was provided.

Upon logging in, each student was shown a screen with their test results. The score sheet for each student contained their own mark (%) in addition to the average mark (%) of the students from the same college, peers at the same level from other colleges, and among all participants. This applied for the test as a whole and for each component. Students could also access their old results and were encouraged to provide feedback.

Thus, each student received their results as follows:

- General overall score compared to relevant batch
- Score in each exam component:
  - Discipline

Grade item	Grade	Rank	Feedback
Progress test results All grades are displayed as (Percentage)			
Discipline wise			
Pharmacotherapy	0.00	1710/3806	The National average is: 4 % And your University Average is: 0 %
Physiology	19.00	1606/3806	The National average is: 16 % And your University Average is: 5 %
Anatomy	17.60	1681/3806	The National average is: 19 % And your University Average is: 10.5%
Biochemistry	0.00	2118/3806	The National average is: 12 % And your University Average is: 21 %
Community	0.00	2827/3806	The National average is: 10 % And your University Average is: 5.5 %
Derma	0.00	1555/3806	The National average is: 6.5% And your University Average is: 4 %
Emergency	0.00	2147/3806	The National average is: 9 % And your University Average is: 4.5 %
ENT	0.00	1499/3806	The National average is: 2.5 % And your University Average is: 1.5 %
Family	0.00	1668/3806	The National average is: 2 % And your University Average is: 0 %
Forensic	5.57	1938/3806	The National average is: 19 % And your University Average is: 19 %
General Surgery	0.00	2618/3806	The National average is: 7 % And your University Average is: 4 %
Histology	14.70	2119/3806	The National average is: 24 % And your University Average is: 10 %
Int Medicine	100.00	2/3806	The National average is: 13.5% And your University Average is: 11.5%
Medical Ethics	0.00	1388/3806	The National average is: 21.5% And your University Average is: 6 %
Micro-immunology	10.00	1675/3806	The National average is: 7 % And your University Average is: 7 %
ObGyn	0.00	2813/3806	The National average is: 13.5% And your University Average is: 12 %
Ophthalmology	17.60	1331/3806	The National average is: 7 % And your University Average is: 2 %
Orthopedics	11.70	1628/3806	The National average is: 5.5 % And your University Average is: 2.5 %
Pathology	15.40	1770/3806	The National average is: 11.5% And your University Average is: 17.5%
Paediatrics	0.00	2537/3806	The National average is: 7.5 % And your University Average is: 9 %
Psychiatry	35.30	668/3806	The National average is: 9.5 % And your University Average is: 5 %
Radiology	11.10	1305/3806	The National average is: 3.5 % And your University Average is: 3 %

Figure 1: Example of the discipline-based report given to students.

- Body systems/areas
- Physician activity
- Dimensions of care
- SaudiMEDs framework themes
- SaudiMEDs PLO

Figures 1 and 2 provide an example of student results and how they appear when downloaded.

*Colleges.* Each participant college was provided with the general results for the test and an analysis including a comparison of its performance to the average performance of all colleges, as well as anonymized comparisons of colleges with regards to the total scores and sub-scores for each component. Each college also received a detailed summary of the results of its own students. For each test item, the college received a comparison of the difficulty index for its students in each batch compared to that of the average at the same program level (Figure 3).

#### Phase 8: Updating the test blueprint

In 2018, a major revision of PT blueprinting was performed to keep pace with Saudi medical education directives.

SaudiMEDs core competencies, initiated by the Saudi Medical Deans Committee, were recently updated and approved by the National Center for Academic Accreditation and Evaluation (NCAAA).<sup>14</sup> Moreover, the Saudi Commission for Health Specialties (SCFHS) circulated the outline of a new blueprint for the Saudi Medical License Examination (SMLE).<sup>16</sup> QCM has revised the PT blueprint in alignment with national milieus.

Three categories were introduced into the existing blueprint: physician activity, dimension of care and SaudiMEDs themes. A physician is expected to be involved in numerous activities: diagnosis, management, communication and professional behavior.<sup>16</sup> In addition, physicians should also be able to provide care in different dimensions: acute, chronic, and psychosocial, as well as support health promotion and disease prevention. The six SaudiMEDs themes—scientific approach to practice, patient care, research and scholarship, communication and collaboration, professionalism and community-oriented practice, were also aligned with the test blueprint.<sup>3,14</sup> The test blueprint was thus revised to incorporate these categories. All test items essentially target common clinical presentations and acute conditions where early intervention is crucial. We believe that the updated PT will enhance our student's capabilities to meet the current and future needs of KSA.

Organ System wise			
<input checked="" type="checkbox"/> Blood and lymph	14.10	2229/3806	The National average is: 19 % And your University Average is: 15.5%
<input checked="" type="checkbox"/> Cardiovascular	0.00	2421/3806	The National average is: 4 % And your University Average is: 3.5 %
<input checked="" type="checkbox"/> Digestive	13.50	1649/3806	The National average is: 8.5 % And your University Average is: 4 %
<input checked="" type="checkbox"/> Hormones–metabolism	0.00	2805/3806	The National average is: 15.5% And your University Average is: 24.5%
<input checked="" type="checkbox"/> Kidney	8.82	1548/3806	The National average is: 5.5 % And your University Average is: 3.5 %
<input checked="" type="checkbox"/> Mental healthcare	0.00	2281/3806	The National average is: 6 % And your University Average is: 2.5 %
<input checked="" type="checkbox"/> Methods–Applied–Biostatistics	4.70	2125/3806	The National average is: 8.5 % And your University Average is: 2 %
<input checked="" type="checkbox"/> Miscellaneous	0.00	2912/3806	The National average is: 14 % And your University Average is: 7.5 %
<input checked="" type="checkbox"/> Molecular –Cellular–Aspects	12.40	1056/3806	The National average is: 6.5% And your University Average is: 6.5 %
<input checked="" type="checkbox"/> Musculoskeletal	17.60	1487/3806	The National average is: 10.5% And your University Average is: 6.5 %
<input checked="" type="checkbox"/> Nervous System–senses	26.40	1092/3806	The National average is: 13.5% And your University Average is: 11.5%
<input checked="" type="checkbox"/> Personal Social Aspects	17.60	1527/3806	The National average is: 15 % And your University Average is: 7.5 %
<input checked="" type="checkbox"/> Reproductive	13.00	1764/3806	The National average is: 10.5% And your University Average is: 3 %
<input checked="" type="checkbox"/> Respiratory	0.00	3076/3806	The National average is: 12.5% And your University Average is: 11 %
<input checked="" type="checkbox"/> Skin	9.50	1586/3806	The National average is: 5.5 % And your University Average is: 4 %
<input checked="" type="checkbox"/> Grand Total	35.30	450/3806	The National average is: 10.5% And your University Average is: 7.5 %

Figure 2: Example of the system-based report given to students.

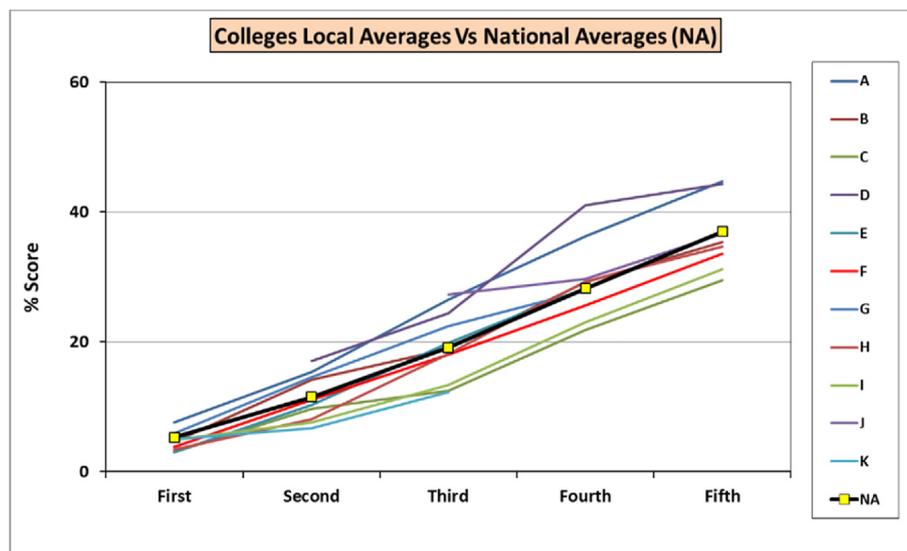


Figure 3: Line graph showing the progress of students at different levels from colleges participating in the 7th PT.

Table 1: Characteristics of colleges and students in different rounds of the PT.

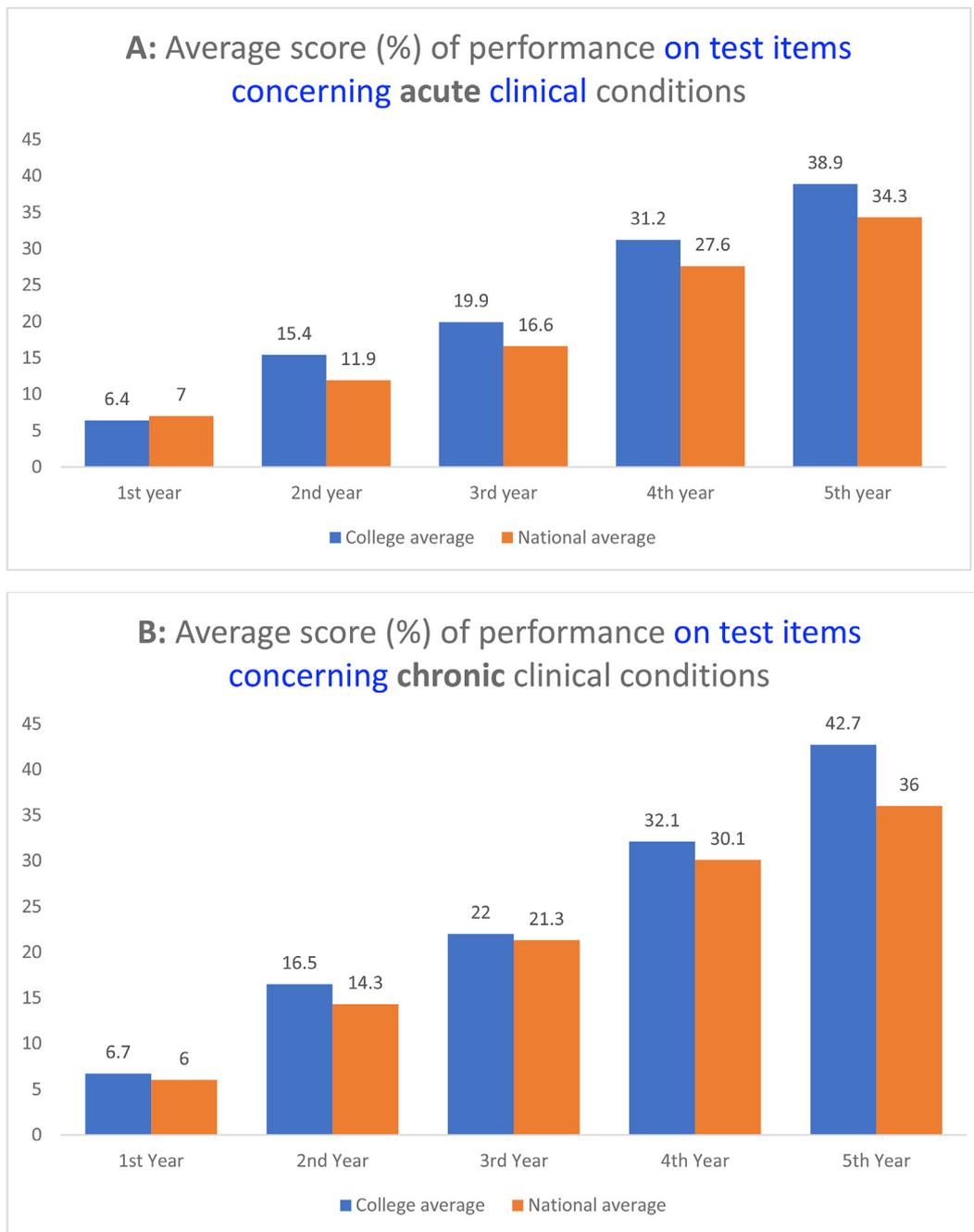
Test round	Month/year	Number of participant colleges			Participant students: n (%)		
		Govern <sup>a</sup>	private	Total	Males, n (%)	Females, n (%)	Total
1	May 2012	3	0	3	240 (84.2%)	45 (15.8%)	285
2	Nov 2012	12	0	12	2122 (55.4%)	1708 (44.6%)	3830
3	April 2013	8	4	12	1403 (54.6%)	1167 (45.4%)	2570
4	Nov 2013	7	3	10	1202 (68.1%)	564 (31.9%)	1766
5	April 2014	8	4	12	1402 (54.7%)	1160 (45.3%)	2562
6	Feb 2015	11	4	15	2720 (64.0%)	1532 (36.0%)	4252
7	Nov 2015	7	4	11	1700 (55.2%)	1377 (44.8%)	3077
8	April 2016	6	2	8	1338 (68.6%)	612 (31.4%)	1950
9	Feb 2017	15	7	22	3419 (55.8%)	2707 (44.2%)	6126
10	March 2018 <sup>b</sup>	17	8	25	4517 (54.5%)	3777 (45.5%)	8294
11	March 2019 <sup>b</sup>	17	7	24	3565 (48.7%)	3748 (51.3%)	7313
12	March 2020 <sup>b</sup>	15	6	21	3481 (50.4%)	3429 (49.6%)	6910
13	March 2021 <sup>b</sup>	15	5	20	4163 (51.2%)	3972 (48.8%)	8135

<sup>a</sup> Govern: Governmental.

<sup>b</sup> Rounds conducted after revision of the blueprint.

Table 2: Overall average marks (%) of students of different year levels in rounds of PT.

Test round	Month/year	Average mark (%) of students at different program levels				
		First year	Second year	Third year	Fourth year	Fifth year
1	May 2012	3.0	13.0	20.5	24.0	43.0
2	Nov 2012	4.0	10.5	16.5	24.2	37.0
3	April 2013	6.7	10.9	18.4	22.5	36.7
4	Nov 2013	3.9	14.1	20.9	32.0	38.6
5	April 2014	6.4	12.8	19.6	30.0	35.3
6	Feb 2015	6.0	11.4	19.2	28.1	34.0
7	Nov 2015	5.3	11.5	19.1	28.2	37.0
8	April 2016	6.4	14.5	18.6	32.0	35.2
9	Feb 2017	6.0	15.1	24.7	32.7	40.6
10	March 2018	7.9	13.2	18.9	31.4	41.3
11	March 2019	5.1	15.1	19.2	27.1	42.6
12	March 2020	7.8	14.9	22.0	30.0	35.8
13	March 2021	7.2	12.6	20.4	30.1	38.5



**Figure 4:** (A & B): Progression of student performance in the PT.

## Results

Throughout the past ten years (May 2012–March 2021), 13 rounds of the PT were conducted. [Table 1](#) shows student participation in the tests. The number of participating colleges increased from three (with 285 participating students) in the first test (May 2012) to more than 20 (with >6000 students) from the ninth round (February 2017) onwards. The participating colleges represented both governmental and private sectors.

In every test, the marks of the students increased as the level of the program increased. The average score (%) of

first-year students ranged from 3.0% to 7.9% and ranged from 34.0% to 43.0% for fifth-year students. [Table 2](#) shows the average marks (%) of students at different levels in different exams.

In each test, there was variation in the average marks of students at different levels of study. [Figure 4](#) shows two examples of the increase in student scores at different levels of the group of colleges participating in the seventh test conducted in November 2015. [Figure 4 \(A\)](#) portrays the performance of students on test items pertaining to acute clinical conditions, while [Figure 4 \(B\)](#) represents their performance on items related to clinical medical conditions.

## Discussion

Since its introduction in the late 1970s,<sup>7</sup> the use of PTs has increased worldwide.<sup>21</sup> In KSA, the QCM is the leading college in preparing and conducting the multi-institutional PT.<sup>6</sup>

The increased number of participating colleges and repetition of participation throughout the ten years indicate the acceptance of the PT modality by most of the colleges. The almost steady increase in participation could be attributed to the appropriate perception of college leaders about the benefits of the test, especially after readjustment of the test blueprints to the SMLE. Changing the policy from two exams per year to only one in response to feedback from participants also contributed to increased participation.

The results of the present study showed that PT can monitor the progression of student knowledge as they advance in their program. The marks of students in each test showed a gradual increase as they advanced in their program. This concurs with earlier results describing the increasing achievement of students as they progressed in their programs.<sup>22</sup>

The results of different tests also showed variation in the rates of progression among different colleges, which was not unexpected and likely resulted from differences in curriculum approaches among colleges; many causes were proposed for this among students at different medical colleges.<sup>22</sup> However, the overall picture shows that the medical programs in KSA provide reasonable cumulative learning through advancement in study.<sup>6</sup>

The results of the present study showed that the average marks at different levels were apparently quite low. Test items were set at the level of graduates and primarily addressed common clinical problems or high-risk situations where early intervention is crucial.<sup>7,10,11,23</sup> These items are based on clinical conditions that require the integration of knowledge from different courses, and students at the early phase of their programs would not be able to respond correctly to most of these questions.

On the other hand, the average score for fifth-year students, ranging between 34% and 43%, may prove that the test items are actually set at the new-graduate level. Based on item-response theory, the chance that an average student responds correctly to an item at their level is 50%. Considering the penalty for wrong answers (−0.25 mark for each wrong response), the average marks of students would be close to 37.5% ( $50 - 0.25 \times 50$ ) if they do not select “do not know.” This would reflect the appropriate construction of test items to target. Furthermore, it has been reported that well-constructed test items are associated with the enforcement of clinical reasoning, problem solving and creative thinking of students.<sup>24</sup>

### *Difficulties encountered and overcome*

It was not easy to reach the present situation in which almost 66% of medical colleges in KSA (27 out of 41) participated in this “optional” test. Difficulties were met within colleges and among participants.

Faculty resistance was the first obstacle, with questions including the following: *What is the philosophy of having a common test for students of all levels since it is expected to have*

*different levels of achievements by students at different levels of the program? How would this test be beneficial for students? Do the expected benefits deserve the time and effort to do it?* These questions were the main bases for objections. The dedicated orientation sessions and workshops with participation from assessment experts from MU and the projection of results for the pilot test were helpful in this regard.

The preparation of *de novo* transdisciplinary test questions linked to the medical curriculum rather than QCM courses and targeting commonly encountered conditions and emergencies where early invention is crucial at the level of new graduates; this represented another challenge. The participation of area experts of all departments in formulating the test blueprint, together with the efforts of the assessment unit in conducting several item-writing workshops, setting criteria for questions to be included in the test, and meticulously reviewing items to ensure they were not used for course exams in the colleges, was critical in overcoming this obstacle.

The leakage of test questions after exams is an additional issue; however, the PT has more benefits than drawbacks, as students will learn from the test questions. Nevertheless, it is a tedious process to prepare new items for each test to ensure validity. This encouraged us to establish a question bank for the PT to ensure the availability of new high-quality items for every exam.

The journey was never smooth; the preparation of questions was a tedious process in order to target medical undergraduate PLOs not specific to a particular college with extreme care for appropriate sampling, undergraduate level and item quality.

Holding the test on the same day for all colleges is another difficulty. This was (and still is) managed by setting an exam date well before the start of planning for the new academic year. The provisional date is communicated to all colleges and their feedback is considered for readjustment when needed.

Incomplete participation of some colleges hinders the maximizing advantages of the test. As the test is optional, some colleges prefer not to participate, while others may select high-achiever students to sit for the test in order to obtain a higher rank. This makes it difficult to achieve accurate benchmarks. This issue could not be overcome by QCM and may need a nation-wide central control mechanism (action).

Transportation time constituted another strain; being a paper and pencil test, necessitated the return of answer sheets to QCM for machine checking and results analysis. This delay has often delayed the declaration of results. For example, for the last exam, the papers of from one college were lost by the courier.

During the lockdown arising due to the COVID-19 pandemic, the 13th round of the PT, held in March 2021, combined online (web-based) as well as on-campus (paper-based) assessments according to the facilities of the participating colleges. Earlier studies showed that there was no significant difference for the students results when compared between online (web-based) and on-campus (paper-based) assessment.<sup>25–27</sup>

### *Factors for success*

We consider that the sustainability of the annual holding of this optional test with increasing numbers of participating

colleges is evidence of the success of the multi-institutional PT venture. It should be emphasized that the continuous support of the Saudi Medical Deans Committee providing an “official-like” national umbrella for the test is a key factor in its success.

The self-motivation of the college and its enthusiastic PT committee helped ensure the quality and relevance of the test items, considering the medical education directives in the Saudi context and adequately responding to feedback from colleges and students (e.g., conducting the test annually rather than twice per year); these are essential elements of test sustainability.

The assessment procedures and the precise analysis of the results guarantee fairness for all participants. Student answers are strictly manipulated and performance is analyzed through a software template created and developed by our team to ensure high quality analysis with many checkpoints in every step.

The method of result declaration is another area of strength. This includes anonymous general results for the colleges in which every college is aware of its position relative to others without knowing the identity of the others, with specific results given to each college about the performance of its students in each component of the test and detailed results given to each student about their performance in each component compared to peers in their own college and throughout KSA. Finally, the adequacy of the logistics maintained the smooth implementation of the different rounds of testing.

#### *Opportunities*

PTs can provide an on-the-go longitudinal assessment tool that monitors student progress at the same time as it exposes pitfalls in the program or its implementation.<sup>28</sup> Our 10-year experience of running of the multi-institutional proved effective as more and more colleges are participating every round. Improvement is both desirable and possible, and the nationalization of the test and online implementation would offer a paradigm shift in the test.

Online/web-based implementation would not only save transportation time, but also the time and human resources necessary for printing exam papers and scanning answer sheets. It would also facilitate the immediate announcement of many of the test results and provide timely feedback to students concerning their performance, including the test items that they considered difficult or unclear. Many studies have shown there to be no difference in performance between paper-based and web-based tests.<sup>25,26,29,30</sup> Importantly, as this is a formative assessment tool, there is a need for strict measures to ensure that no misconduct occurs.

The nationalization of the test will provide significant advantages: it would aid in mandating of the test to include all medical colleges in KSA, thus providing them with knowledge about the positions of their students among peers throughout their learning journey, including areas requiring improvement and corrective actions before graduation. Some colleges do not allow final-year students to sit for the end-of-program exam unless they have reached a specific threshold in the PT.<sup>22</sup> It is clear that nationalization of the test requires a central authorized supervision body and a consortium, rather than a

single college, to conduct the test. The consortium would increase the human resources needed for item writing, thus increasing the efficiency of test items. The central authorized body will be able to oversee the benchmarking of the participating colleges, thus identifying areas for improvement and providing an opportunity to provide advice for correcting actions when needed, thus assuring confidentiality. Thus, nationalization will upgrade the test quality and benefit students through its implementation, thus leading to the further improvement of medical education with increasing competence of graduates and improvement of the healthcare system in KSA; this is one of the issues in the Saudi vision 2030 and the national transformation program.<sup>31</sup>

#### *Limitations*

The authors realize that it would have been a very strong indicator of the success of the PT if we could correlate students marks in the PT to their performance in the Saudi Medical Licensing Exam (SMLE). In order to do this, we would need to access the results of SMLE for different participants; we believe that this may be done individually for each college to retain confidentiality. Another limitation is the lack of information about how participating colleges might use the results of the PT as a diagnostic tool for continuous improvement of the curriculum.

Finally, there are no inferential statistics or psychometric properties of PT items in the data presented here. Although these data are essential, they are beyond the scope of the present study; therefore, the College intends to conduct a comprehensive retrospective study (including all previous tests) in order to present these data and emphasize their implication.

#### **Conclusions**

The Saudi PT is remarkable for several reasons. The PT is a curriculum-independent test covering the end level of medical curricula in KSA and provides constructive feedback for test takers and other stakeholders about their positions on the national level. The PT is a meticulously crafted test to evaluate the essential knowledge of a medical graduate. As it is considered a robust assessment tool of cognitive skills, it was widely accepted by participating colleges.

Further feedback from colleges and students would be helpful for continuous improvement. Moreover, the opinions of graduates with regards to the s of progress testing with regards to their performance in the license exam and real practice should be sought. These issues should be investigated in future research.

In addition, the correlation of scores in the PT and SMLE, and the impact of the PT on the continuous improvement of curricula is a good platform for nation-wide multi-institutional research.

#### **Source of funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not for profit sectors.

## Conflict of interest

The authors have no conflict of interest to declare.

## Ethics approval

Because this research was carried out in accordance with the principles of the Declaration of Helsinki, the local ethics committee at Qassim University's College of Medicine determined that it did not require formal ethics approval. Furthermore, because the study is based on data collected during test implementation, is completely anonymous, and does not involve direct contact with human subjects, written informed consent was not required.

## Authors contributions

ASA, MNEDS, and MMW initiated the research idea and wrote and edited the manuscript. TARS entered, analyzed and tabulated the data. HA, AA, and HA finalized the manuscript, inserted citations, wrote the abstract and introduction, and prepared the manuscript for publication. All authors have read and approved this manuscript. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

## References

- Telmesani A, Zaini R, Ghazi H. Medical education in Saudi Arabia: a review of recent developments and future challenges. *EMHJ-East Medit Health J* 2011; 17(8): 703–707.
- Alamro AS. Analysing undergraduate medical curricula: experience from a Saudi medical college. *Majmaah J Health Sci* 2019; 7(3): 20–33.
- Zaini RG, Bin Abdulrahman KA, Al-Khotani AA, Al-Hayani AMA, Al-Alwan IA, Jastaniah SD. Saudi Meds: a competence specification for Saudi medical graduates. *Med Teach* 2011; 33(7): 582–584.
- Tekian A, Almazrooa AA. Does Saudi Arabia need an Abraham Flexner? *Med Teach* 2011; 33(1): 72–73. <https://doi.org/10.3109/0142159X.2010.528475>.
- Zaini RG. National consensus of the vision of the " Saudi future doctor. In: *Current status and future perspective of medical education in Saudi medical schools*. University of Sheffield; 2007.
- Al-Shobaili H, Nor-El-Din M, AlGhasham A, Saqr M. *Report of the first multi-institutional progress test in Saudi Arabia*. Prague, Czech Republic: AMEE Conference Prague Congress Centre; 2013 [cited 2019 May] Available from: <https://amee.org/getattachment/Conferences/AMEEPast-Conferences/AMEE-Conference-2013/AMEE-2013-ABSTRACT-BOOK-updated-190813pdf2013>.
- Vleuten CVD, Verwijnen G, Wijnen W. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach* 1996; 18(2): 103–109.
- Blake JM, Norman GR, Keane DR, Mueller CB. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med* 1996; 71(9): 1002–1007. <https://doi.org/10.1097/00001888-199609000-00016>.
- Nouns ZM, Georg W. Progress testing in German speaking countries. *Med Teach* 2010; 32(6): 467–470. <https://doi.org/10.3109/0142159X.2010.485656>.
- Schuwirth LW, van der Vleuten CP. The use of progress testing. *Perspect Med Educ* 2012; 1(1): 24–30.
- Wrigley W, Van Der Vleuten CP, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Med Teach* 2012; 34(9): 683–697.
- Pugh D, Regehr G. Taking the sting out of assessment: is there a role for progress testing? *Med Educ* 2016; 50(7): 721–729. <https://doi.org/10.1111/medu.12985>.
- Alrebish SA, Taha MH, Ahmed MH, Abdalla ME. Commitment towards a better future for medical education in Saudi Arabia: the efforts of the college of medicine at Qassim University to become socially accountable. *Med Educ Online* 2020; 25(1):1710328. <https://doi.org/10.1080/10872981.2019.1710328>.
- NCAAA. SaudiMEDs framework. Available at: <https://www.ncaaa.org.sa/enportal/Accreditation/Programmatic/pages/medical-colleges.aspx> 2017; 2017.
- Tekian AS, Al Ahwal MS. Aligning the SaudiMED framework with the national commission for academic accreditation and assessment domains. *Saudi Med J* 2015; 36(12): 1496–1497. <https://doi.org/10.15537/smj.2015.12.12916>.
- SCFHS. Saudi medical licensing exam (SMLE) guidelines. Available at:<https://www.scfhs.org.sa/en/examinations/Exams/Blueprints/Pages/default.aspx#tab-92-22017>.
- Verhoeven B, Verwijnen G, Scherpbier A, Schuwirth L. Quality assurance in test construction: the approach of a multidisciplinary central test committee/Commentary. *Educ Health* 1999; 12(1): 49.
- Ravesloot C, Van der Schaaf M, Muijtjens A, Haaring C, Kruitwagen C, Beek F, et al. The don't know option in progress testing. *Adv Health Sci Educ* 2015; 20(5): 1325–1338.
- Tio RA, Schutte B, Meiboom AA, Greidanus J, Dubois EA, Bremers AJ. The progress test of medicine: the Dutch experience. *Perspect Med Educ* 2016; 5(1): 51–55.
- Hv Mameren, Cvd Vleuten. The effect of a 'don't know'option on test scores: number-right and formula scoring compared. *Med Educ* 1999; 33(4): 267–275.
- Freeman A, Van Der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach* 2010; 32(6): 451–455.
- Albanese M, Case SM. Progress testing: critical analysis and suggested practices. *Adv Health Sci Educ* 2016; 21(1): 221–234.
- Al-Shobaili H, Saleh MN-E, AlGhasham A, Saqr M. *Report of the first multi-institutional progress test in Saudi Arabia*. Prague, Czech Republic: AMEE Conference; 2013. Available at: <https://amee.org/getattachment/Conferences/AMEE-Past-Conferences/AMEE-Conference-2013/AMEE-2013-ABSTRACT-BOOK-updated-190813.pdf>. [Accessed May 2017]. Accessed on.
- Dugdale AE. The Pathway MCQ: a method for teaching and testing deeper knowledge. *Med Teach* 1998; 20(3): 250–253. <https://doi.org/10.1080/01421599881011>.
- Ford BD, Vitelli R, Stuckless N. The effects of computer versus paper-and-pencil administration on measures of anger and revenge with an inmate population. *Comput Hum Behav* 1996; 12(1): 159–166. [https://doi.org/10.1016/0747-5632\(95\)00026-7](https://doi.org/10.1016/0747-5632(95)00026-7).
- Wang S, Jiao H, Young MJ, Brooks T, Olson J. Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: a meta-analysis of testing mode effects. *Educ Psychol Meas* 2008; 68(1): 5–24. <https://doi.org/10.1177/0013164407305592>.
- Saleh MN-E-D, Salem TARo, Alamro AS, Wadi MM. Web-based and paper-based examinations: lessons learnt during the COVID-19 pandemic lockdown. *J Taibah Univ Med Sci* 2022; 17(1): 128–136. <https://doi.org/10.1016/j.jtumed.2021.09.004>.
- Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessments: feedback for individuals and institutions based on the progress test. *Med Teach* 2010; 32(6): 486–490. <https://doi.org/10.3109/0142159X.2010.485652>.

29. Boevé AJ, Meijer RR, Albers CJ, Beetsma Y, Bosker RJ. Introducing computer-based testing in high-stakes exams in higher education: results of a field experiment. *PLoS One* 2015; 10(12):e0143616. <https://doi.org/10.1371/journal.pone.0143616>.
30. Saleh MN-E-D, Salem TARo, Alamro AS, Wadi MM. Web-based and paper-based examinations: lessons learnt during the COVID-19 pandemic lockdown. *J Taibah Univ Med Sci* 2021. <https://doi.org/10.1016/j.jtumed.2021.09.004>.
31. Albejaidi F, Nair KS. Building the health workforce: Saudi Arabia's challenges in achieving Vision 2030. *Int J Health Plann*

**Manag** 2019; 34(4): e1405–e1416. <https://doi.org/10.1002/hpm.2861>.

**How to cite this article:** Alamro AS, Alghasham AA, Al-Shobaili HA, Alhomaïdan HT, Salem TA, Wadi MM, Saleh MN. 10 years of experience in adopting, implementing and evaluating progress testing for Saudi medical students. *J Taibah Univ Med Sc* 2023;18(1):175–185.