

**Original article:**

**Progress Testing and Module Final Scoring as part of Curriculum Evaluation in Faculty of Medicine**

**Syarif Hidayatullah State Islamic University Jakarta, Indonesia**

*Marita Fadhillah<sup>1</sup>, Nurmila Sari<sup>1</sup>, Sophie Dwiyantri<sup>1</sup>, Erike A. Suwarsono<sup>1</sup>, Fika Ekayanti<sup>1</sup>*

**Abstract**

**Introduction:** Progress testing (PT) reflects students' knowledge development and is a valuable indicator for curriculum evaluation. Since 2009, Faculty of Medicine Syarif Hidayatullah State Islamic University Jakarta (FMSH) has been conducting PT every year as a formative assessment. In 2012, curriculum reform was addressed to revise the 2005 curriculum; until then PT and curriculum evaluation were not conducted concurrently. This study aims to evaluate PT and assess whether there is a relationship between PT performance and final scores in modules, as part of curriculum evaluation. **Method:** It reviews PT for two cohorts: 571 students in 2015 and 562 students in 2016. 120 system-based topics were addressed in the PT. In this study the final scores for the old (2015) and new (2016) curriculum neuropsychiatry modules are reviewed, since their scores were lower than for other modules. Comparisons were made using ANOVA. Pearson correlations were calculated to examine the relationship between PT and final module scores. **Results:** This study revealed that PT scores between each grade ( $p < 0.001$ ) from 2015 to 2016 improved significantly ( $54.49 \pm 7.43$  and  $55.07 \pm 8.32$ ;  $p < 0.001$ ). The mean of the final score of the new neuropsychiatry module was  $69.36 \pm 3.78$  while the old one was  $70.92 \pm 3.99$ . Furthermore, Pearson correlation showed a weak correlation between final scores for the neuropsychiatry module and PT scores in 2015 ( $\rho = 0.191$ ,  $p = 0.011$ ). **Discussion:** PT scores increased significantly. Despite the final score of the new neuropsychiatry module being lower than the old one, there was heterogeneity in scores within the old neuropsychiatry module. The small number of neuropsychiatry items in the PT explains why the correlation between PT and final scores was weak. The weak correlation between final scores for the neuropsychiatry module and the PT scores in 2015; PT and final module scores seem reliable as indicators of curriculum evaluation. Further study is needed to analyze more cohort PT scores and modules.

**Keyword:** Progress test, Module score, Module evaluation, Curriculum evaluation, Curriculum reform

*International Journal of Human and Health Sciences Vol. 05 No. 01 January'21 Page : 62-68  
DOI: <http://dx.doi.org/10.31344/ijhhs.v5i1.235>*

**Introduction:**

Progress testing (PT) has become widely used in many medical and health institutions all over the world,<sup>1-10</sup> since it was introduced for the first time in 1970 by the University of Missouri-Kansas City School of Medicine and Maastricht University in Netherlands.<sup>11</sup> PT's growth is related to its characteristics in providing longitudinal and repeat assessment for students during their training. In addition, implementing PT affects a change in thinking about assessment, as well as change in

academic culture. In general, PT is a longitudinal-test approach based on equivalent tests given at fixed intervals, with the intention of assessing the development of functional knowledge and/or competence.<sup>12</sup> Indeed, PT predicts students' future competence and/or performance and can therefore reduce failure rates in the licensing examination.<sup>13</sup> As problem-based learning (PBL) approaches are implemented in most medical schools worldwide, it is recommended that, as part of PBL approaches, schools should use PT in order to assess the

1. Faculty of Medicine, Syarif Hidayatullah State Islamic University Jakarta, Indonesia

**Correspondence to:** Marita Fadhillah, MD, Ph.D, Faculty of Medicine, Syarif Hidayatullah State Islamic University Jakarta, Indonesia. **Email:** [maritafadhillah@uinjkt.ac.id](mailto:maritafadhillah@uinjkt.ac.id)

progress of both students and the curriculum.<sup>11,14,15</sup> PT not only assesses functional knowledge or competence in basic medical science,<sup>16</sup> but also helps students of clinical science to develop their higher cognitive processes.<sup>14,17-18</sup>

PT can also encourage students' clinical reasoning processes,<sup>19</sup> and recently many institutions have been developing PT as an assessment tool for students' clinical skills. The other advantages of PT is that it provides individual and institutional feedback.<sup>20</sup>

In terms of PT implementation, cross-institutional strategies are now commonly implemented to reassure students, faculty, and accreditation bodies of the comparative level of knowledge acquisition between medical schools.<sup>3</sup> In terms of these benefits, cross-institutional PT can offer openness, non-competitiveness, exchange, and mutual trust.<sup>12</sup> Some countries, including the Netherlands, United Kingdom, United States of America, Germany, Austria, Saudi Arabia, and Indonesia, have already conducted cross-institutional or collaborative PT.<sup>3,6,9,21</sup>

Since 2009, we have been conducting PT once a year as formative assessment. It provides feedback to students about how far they have progressed and gets them used to the national exit examination. It is also a learning-process evaluation for the faculty.

When Faculty of Medicine Syarif Hidayatullah State Islamic University Jakarta (FMSH) was established in 2005, its curriculum was adopted from Faculty of Medicine University of Indonesia (FMUI) and was followed until 2011. Then in 2012 curriculum reform was achieved and we moved from the 2005 to the 2012 curriculum. The main difference between these curricula is the focus; in the new approach, integrated physiology modules in the first year are followed by pathology modules in the next year, while the previous curriculum used a pathophysiology approach from the first year.

Until now, we have not carried out both PT and curriculum evaluation. As a developing medical school, it is vital that students and faculty are assured by evidence that the learning process in the new curriculum is adequate. For these reasons, at the end of the first decade of our medical school we would like to:

1. Evaluate PT through describing the evaluation of PT in 2015 and 2016 and assessing whether there is a difference between PT results in 2015 and 2016.
2. Assess whether there is a correlation between PT and final module scores as part of our new curriculum evaluation.

## Educational Context

### Progress testing in Faculty of Medicine Syarif Hidayatullah State Islamic University Jakarta (FMSH)

FMSH is one of Indonesia's younger medical schools, having been established in 2005. Until 2011, we were supported by FMUI. The FMSH curriculum was adapted from FMUI's curriculum, which is system based, student centered, and community oriented, with additional local content appropriate for Muslim doctors. The first PT in FMSH was conducted in 2009, the blueprint for which came from FMUI. Since 2011, we have been administering PT independently.

PT in FMSH is conducted as formative assessment once a year, usually in the middle of the academic year. PT has some specific objectives that not only provide feedback for students but also for the faculty/institution.<sup>18</sup> Students can assess their own knowledge ability over time, and become familiar with the national exit examination model, which requires higher cognitive abilities. PT also provides feedback to faculty/institution about learning-process evaluation.

All students of FMSH, from first to fifth grades, and from preclinical to clinical phases, are enrolled in the PT program. We inform all students at the beginning of the academic year about the objectives and procedures of PT through direct information and poster announcements.

We have developed a PT blueprint that is based on the Competency Standards for Indonesian Doctors 2012 (locally known as *Standard Kompetensi Dokter Indonesia 2012*). Similarly to the national exit examination, we develop system- and competency-based vignette questions.<sup>22</sup> Ten systems and four additional issues are addressed in the test (Table 1). In terms of the Competency Standards for Indonesian Doctors 2012, we create items that require analytical thinking processes rather than recall, and which cover doctor competency levels 3 or 4, as explained below:

Level of competency 3a: doctor capable of making clinical diagnoses based on physical and additional examination, such as laboratory examination, X-ray, etc. Doctor can make a decision and give early treatment, then refer to relevant specialist (non-emergency case).

Level of competency 3b: doctor capable of making clinical diagnosis based on physical and additional examinations, such as laboratory examination, X-ray, etc. Doctor can make a decision and give early treatment, then refer to relevant specialist (emergency case).

Level of competency 4: doctor capable of making clinical diagnosis based on physical and additional examinations, such as laboratory examination, X-ray, etc. Doctor can make a decision and treat the patient's problems completely and independently. We also include a few questions that cover doctor competency levels 1 or 2, as explained below:

Level of competency 1: doctor capable of detecting and defining clinical conditions that are appropriate to the disease when reading literature. The doctor then knows how to obtain further information. If the doctor meets a patient with the clinical condition, they can assume diagnosis and refer the patient immediately.

Level of competency 2: doctor capable of making clinical diagnosis based on physical and additional examinations, such as laboratory examination, X-ray, etc. Doctor capable of referring the patient immediately to the relevant specialist and can follow up on the patient.

Once the PT blueprint was established, we commenced an internal review by medical education unit staff, supported by expert content resources. This process was very important as it provided process and content-quality control. In this review, we included effective analytical items, and removed poor ones. Finally, we had included 120 analytical items using single best answers from multiple-choice questions (MCQs). We conducted the paper-based PT with a time allowance of two hours.

**Table 1:** Blueprint of system- and competency-based PT

Systems	Number of items (%)	Competencies (% numbers of items)
Reproduction	10 (8.33)	Level 1 (5.83)
Gastroenterology and hepatology	10 (8.33)	
Respiratory	10 (8.33)	
Endocrine and metabolism	10 (8.33)	
Renal and urinary tract	10 (8.33)	
Cardiovascular	10 (8.33)	Level 2 (20.00)
Dermatomusculoskeletal	10 (8.33)	Level 3a (13.33)
Hematoimmunology	10 (8.33)	
Head and neck	10 (8.33)	Level 3b (14.16)
Neuropsychiatry	10 (8.33)	
Community medicine,	10 (8.33)	Level 4 (46.67)
Basic research	3 (2.50)	
Bioethics	4 (3.33)	
Communication and professionalism	3 (2.50)	
Total	120 (100)	

To obtain valid and reliable results, we used an item-analysis and absolute-standard-setting approach. Finally, we announced the result of PT for each grade.

#### **Curriculum reform in Faculty of Medicine Syarif Hidayatullah State Islamic University Jakarta (FMSH)**

When FMSH was established in 2005, a PBL approach was applied as the main learning method for students. In the beginning, we adopted the curriculum of FMUI, since we were supported by them. Some local content was added to enrich learning for medical students with Islamic values. This curriculum is known as the old curriculum or the 2005 curriculum.

In 2011, a need assessment of the old curriculum was conducted using the Health Profession Educational Quality project from the Indonesia Ministry of Research, Technology, and Higher Education. Its results showed that mastery by students of the basic medical science that underlies various clinical problems was low. In addition, students faced difficulties in integrating basic medical science and clinical medicine. Curriculum reform was therefore conducted in 2012, and the new curriculum is known as the 2012 curriculum. In the 2005 curriculum, the physiology system was blended into modules along with clinical cases. It was given at the beginning of modules. Thus, we introduced the physiology process in the first week of the module and then gradually increased to complex case pathology in the remaining weeks of the module. The 2005 curriculum represented a spiral principle within modules.

In contrast, the 2012 curriculum starts with physiology system modules during the first year to provide a strong foundation for students and a better understanding of clinical problems. In the following year, students have exposure to pathology in complex systems modules. This is consistent with a spiral principle in the curriculum. Significant difference for the 2012 curriculum to the 2005 curriculum is the delivery method for integration of the module, critical thinking endorsement, and experts are more involved in lecture and students' presentation.

In this study, we used the final scores of six-week neuropsychiatry modules delivered to second year students in the 2012 curriculum and third year students in the 2005 curriculum, thus covering the period of curriculum reform. The final scores contain summative and process scores (such as small group discussion scores, laboratory quiz scores, etc).

#### Methods and Materials:

This study used PT for two-year cohorts: 2015 and 2016. We encouraged all students to participate in PT. In total, 571 students from first to fifth grade were enrolled in PT in 2015 and 562 students in 2016. The PT was composed of 120 single-best-answer MCQs covering ten systems and four additional competencies. It was a paper-based test and lasted approximately two hours.

We also used final score data for the neuropsychiatry modules of the 2005 and 2012 curriculums. In the 2012 curriculum, integrated physiological systems were emphasized in the first year, before pathological systems were introduced in the second year. In the 2005 curriculum, pathological-system-based learning was conducted including physiological systems in early modules, from the first year.

For the study, statistical analysis was addressed using one-way ANOVA with post hoc Bonferroni test evaluation to analyze the progress of students' levels of knowledge. To calculate the correlation between PT and module final scores, we used Pearson correlation.

#### Results:

A total of 571 students in 2015 and 562 students in 2016 were enrolled in the PT reviewed in this study. Table 2 shows the number of students that participated in PT in 2015 and 2016, from the first to sixth year. Where there are no participants, this is because when PT of 2015 was conducted, the 2016 class had not yet enrolled, and when PT of 2016 was conducted, the class of 2009 had already graduated.

**Table 2:** Student participation in PT of 2015 and 2016 (N)

Student class of year	PT of 2015	PT of 2016
Class of 2009	40	0 (has already graduated)
Class of 2010	83	9
Class of 2011	87	95
Class of 2012	89	93
Class of 2013	76	87
Class of 2014	86	92
Class of 2015	110	102
Class of 2016	0 (has not enrolled yet)	84
Total	571	562

Table 3 shows the mean score of PT in 2015 and 2016 from first to sixth grade. The results indicate that there was an increase in PT mean score among the medical students of FMSH, both in 2015 and 2016. In the class of 2010, there was a decrease in mean score of PT from 2015 to 2016. This is because most of the class of 2010 had already graduated and just a few low performance students remained.

Table 3 also shows the minimum and maximum scores for PT for 2015 and 2016 among participating students. Both minimum and maximum scores of PT improved from 2015 to 2016, except the minimum score for the class of 2013 and the maximum score for the class of 2010. This study reveals that the PT scores between each grade were significantly different, and that from 2015 to 2016 they increased significantly ( $F(7, 1124) = 186.429, p < 0.001$ ). The mean scores of PT difference between 2015 and 2016 were  $54.49 \pm 7.43$  to  $55.07 \pm 8.32$ . The post hoc Bonferroni multiple comparison test of PT score is shown in Table 4. This test shows that PT results for younger classes (classes of 2016, 2015, 2014 and 2013) almost all had significant differences to others classes compared to older classes (classes of 2012, 2011, 2010 and 2009).

This study found that the mean of the final score of the new neuropsychiatry module was  $69.36 \pm 3.78$  (CI 95% 68.60–70.13) while for the old module it was  $70.92 \pm 3.99$  (CI 95% 70.11–71.72). Figure 1 shows that in the old neuropsychiatry module there were more outliers than in the new neuropsychiatry module.

In addition, Pearson correlation shows weak correlation between the final score of the neuropsychiatry module and the PT score in 2015

( $p = 0.191$ ,  $p = 0.011$ ). Pearson correlation between the final score of the neuropsychiatry module and the PT score in 2016 was  $p = 0.052$ ,  $p = 0.481$ .

### Discussion and Conclusions:

After six years of carrying out PT and a decade of the establishment of FMSH, the early evaluation of PT and a small part of the curriculum has been conducted. This study provided both positive and negative results, as discussed below.

This study revealed that there was a significant difference in PT scores between each grade, and that from 2015 to 2016 it increased significantly. It was predicted that in the younger class the score would be lower than in the older class. Naturally, the PT score of preclinical students was lower than clinical students, as we performed vignette question and mostly level 4 competency testing in our PT. Usually, preclinical students answer questions using lower-order cognitive processing and clinical students answer questions using higher-order cognitive processing.<sup>17</sup>

This study provided evidence that PT was one of the assessment tools that can show students' progress in functional knowledge during their education in a PBL setting, and that this is so not only for high-performance students but also for those with lower performance. It was a predictable advantage that PT provides constant repetition of topics, thus encouraging long-term and functional-knowledge retention.<sup>11,21</sup> PT can also provide early detection of students with high performance.<sup>11</sup>

The increasing PT score among FMSH students from 2015 to 2016 might relate to the PBL approach that we have been using since 2005. PBL students have better long-term knowledge retention than traditional curriculum students. It seems that discussing knowledge in small groups and in the context of patient problems during the Bachelor's programme can result in better knowledge retention at the end of the Master's programme.<sup>15</sup> In order to keep up with all levels of students and also for curriculum progress in both basic and clinical sciences, it is recommended that PBL schools use PT.<sup>14</sup>

According to multiple comparison by post hoc Bonferroni testing, it appears that the younger classes have made more progress compared to older classes. If we explore growth of knowledge, van Diest mentioned that growth of knowledge in psychiatry and behavioral sciences increased in early years and then leveled off after years 3 and 4, and finally dropped in the last two years.<sup>23</sup>

Although the final score of the new

neuropsychiatry module score being lower than for the old neuropsychiatry module, in the old neuropsychiatry module there was a heterogeneity score. It showed that students' knowledge in the old neuropsychiatry module was varied and further study should be encouraged to identify in details whether prior knowledge or pathological case exposures have impact or not. It seemed many factors could relate to this situation. Since the content delivery and learning process were different between the old and new module, further exploration on those aspects should be consider for future study. In addition, the high score is not the only purpose of assessment, it is more important to improve the low performance students. Since we noticed heterogeneity in the old neuropsychiatry module, it revealed that the old neuropsychiatry module has not improved the low performance students.

The small number of neuropsychiatry items in PT seemed to affect the weak correlation between PT score and the final score of the neuropsychiatry module. Van der Vleuten said in his review that PT reflected the end objective of students' knowledge, thus any change in the curriculum has no direct consequences for organizing PT.<sup>11</sup> In addition, Al Alwan mentions that PT results in behavioral sciences were lower than in basic medical sciences.<sup>14</sup> This could explain why the correlation between PT score and final score of neuropsychiatry module was weak. While these reasons seem to support the weak correlation between PT and final scores for the module, this supports the view that PT is useful for the evaluation of knowledge retention in specific curriculums: in our case, the neuropsychiatry module.<sup>5</sup>

This study provides the first evidence for PT and a small part of curriculum evaluation after a decade of the establishment of FMSH. It provides promising evidence for PT and curriculum evaluation for new medical schools in Indonesia that still lack resources.

Since in Indonesia not all medical schools conduct PT this study provides a success story for PT in a new medical school.

Plessas<sup>6</sup> states that in PT there is a considerable variation in the number of items included (100 to 250 items) and in the frequency of test administration (two to four times per year). Until now, because of resource limitations, we perform PT just once a year with 120 items tested by MCQs. This could under-represent the content of the curriculum, thus



challenging the content validity of the test, and the number of tests conducted in a year may reduce the total of sampling opportunities and hence validity. Another limitation of this study is that we analyzed the correlation only for neuropsychiatry modules.

In this study, we have not explored in detail the reasons why the neuropsychiatry module score that represented the new curriculum was lower than the old one, or why there were more outliers in the old neuropsychiatry module than new one. We will explore this in future studies.

In future study we will also explore more cohorts' PT scores and analyze test items from aspects such as whether they are system-based, basic clinical science, or level-of-competency based.

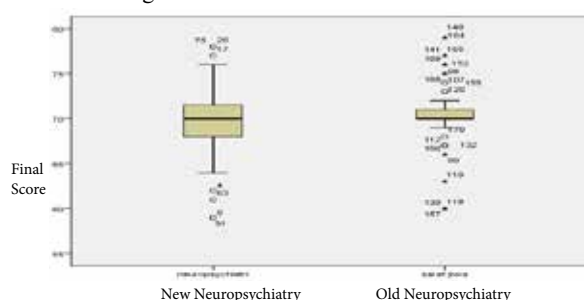
**Table 3:** Mean and minimum-maximum scores of PT results for 2015 and 2016

	Mean score of PT 2015	Minimum and maximum score of PT 2015	Mean score of PT 2016	Minimum and maximum score of PT 2015
Class of 2009	53.20±8.03	37.33-66.67	-	-
Class of 2010	60.85±7.39	41.33-72.00	50.19±13.57	46.00-62.67
Class of 2011	58.47±7.44	42.00-74.67	64.48±8.75	41.33-78.67
Class of 2012	59.68±8.44	35.33-72.00	63.95±8.32	41.33-76.00
Class of 2013	58.28±6.40	46.00-70.67	59.26±7.51	42.67-70.00
Class of 2014	50.68±6.78	34.00-62.67	55.91±6.71	37.33-65.33
Class of 2015	40.32±7.55	32.00-47.33	50.15±6.83	33.33-59.33
Class of 2016	-	-	41.58±6.55	26.00-50.00
Total average	54.50±7.43		55.07±8.32	

**Table 4:** Multiple comparison results of PT scores using post hoc Bonferroni test

	Class of 2009	Class of 2010	Class of 2011	Class of 2012	Class of 2013	Class of 2014	Class of 2015	Class of 2016
Class of 2009	-	<0.001	<0.001	<0.001	<0.001	NS	<0.001	<0.001
Class of 2010	<0.001	-	NS	NS	NS	<0.001	<0.001	<0.001
Class of 2011	<0.001	NS	-	NS	0.004	<0.001	<0.001	<0.001
Class of 2012	<0.001	NS	NS	-	0.004	<0.001	<0.001	<0.001
Class of 2013	<0.001	NS	0.004	<0.001	-	<0.001	<0.001	<0.001
Class of 2014	NS	<0.001	<0.001	<0.001	<0.001	-	<0.001	<0.001
Class of 2015	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	-	0.004
Class of 2016	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.004	-

\*NS = not significant



**Figure 1:** Plot diagram of final scores for new and old neuropsychiatry modules

Since students' satisfaction relates to PT scores,<sup>24</sup> we need to encourage students to reflect on their

perceptions of PT, and also the impact of PT during their education. In addition, we need to explore whether PT scores correlate to students' perceptions of PT. To challenge FMSH students and faculty, we could conduct cross-institutional or collaborative PT with other medical institutions in Indonesia.

This study shows that all grades of students of FMSH made progress significantly between the PT of 2015 and 2016. As part of the new curriculum evaluation, the final scores for the new neuropsychiatry module were lower than for the old one, but there was heterogeneity of scores within the old neuropsychiatry module that we have to explore in future studies. Regarding the weak correlation between the final scores of the neuropsychiatry module and PT scores in 2015, this might relate to the small number of neuropsychiatry items in the PT. PT and final module scores seem reliable as indicators of curriculum evaluation.

#### Ethical Approval:

This study was accepted by the Ethics Committee of Faculty of Medicine and Health Sciences of Syarif Hidayatullah State Islamic University Jakarta, Indonesia.

#### Conflict of interest:

The authors declare that there were no conflicts of interest.

#### Acknowledgement:

The authors would like to acknowledge support from Umatul Khoiriyah, MD, Ph.D for her valuable review and the Institution for Research and Community Engagement, Syarif Hidayatullah State Islamic University Jakarta for its support

through the 2017 Conference Travel Grant.

#### Author's Contributions:

Data gathering and idea owner of this study: Marita Fadhillah

Study design: Marita Fadhillah

Data gathering: Marita Fadhillah, Nurmila Sari and Sophie Dwiyantri

Data analysis and consultation: Marita Fadhillah, Erike A. Suwarsono, and Fika Ekayanti

Writing and submitting manuscript: Marita Fadhillah

## References:

1. Aarts R, Steidell K, Manuel BAF, Driessen EW. Progress testing in resource-poor countries: a case from Mozambique. *Medical Teacher*. 2010; 32:461–463.
2. Wrigley W, van der Vleuten CPM, Freeman A, Muijtjens A. A systematic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Medical Teacher*. 2012; 34:683–697.
3. Soliman MM, Al-Shaikh GK, Alnassar SA. Use of cross-institutional progress test as a predictor of performance in a new medical college. *Advances in Medical Education and Practice*. 2016; 7:197–200.
4. Lillis S, Yelder J, Mogo V, O'Connor B, Bacal K, Booth R, Bagg W. Progress testing for medical students at the University of Auckland: results from the first year of assessments. *Journal of Medical Education and Curricular Development*. 2014; 1:41–45.
5. Schmidmaier R, Holzer M, Angstwurm M, Nouns Z, Reincke M, Fischer MR. Using the progress test Medizin (PTM) for evaluation of the Medical Curriculum Munich (MeCuM). *GMS Zeitschrift für Medizinische Ausbildung*. 2010; 27(5):1–14.
6. Plessas A. Validity of progress testing in healthcare education. *International Journal of Humanities Social Sciences and Education*. 2015; 2(8):23–33.
7. Tio RA, Schutte B, Meiboom AA, Greidanus J, Dubois EA, Bremers AJA, . The progress test of medicine: the Dutch experience. *Perspectives on Medical Education*. 2016; 5:51–55.
8. Tomic ER, Martins MA, Lotufo PA, Benseñor IM. Progress testing: evaluation of four years of application in the School of Medicine, University of São Paulo. *Clinics*. 2005; 60(5):389–396.
9. Findyartini A, Werdhani RA, Iryani D, Rini EA, Kusumawati R, Poncorini E, Primaningtyas W. Collaborative progress test (cPT) in three medical schools in Indonesia: the validity, reliability and its use as a curriculum evaluation tool. *Medical Teacher*. 2015; 37(4):366–373.
10. Matsuyama Y, Muijtjens AMM, Kikukawa M, Stalmeijer R, Murakami R, Ishikawa S, Okazaki H. A first report of East Asian students' perception of progress testing: a focus group study. *BMC Medical Education*. 2016; 16:245.
11. Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher*. 1996; 18:103–110.
12. Schuwirth LWT, van der Vleuten CPM. The use of progress testing. *Perspectives on Medical Education*. 2012; 1:24–30.
13. Norman G, Neville A, Blake JM, Mueller B. Assessment steers learning down the right road: impact of progress testing on licencing examination performance. *Medical Teacher*. 2010; 32(6):496–499.
14. Al Alwan I, Al-Moamary M, Al-Attas N, Al Kushi A, AlBanyan E, Zamakhshary M, Al Kadri HMF, Tamim H, Magzoub M, Hajeer A, Schmidt H. The progress test as a diagnostic tool for a new PBL curriculum. *Education for Health*. 2011; 24(3):1–10. Available from [http://www.educationforhealth.net/temp/EducHealth243493-8598473\\_235304.pdf](http://www.educationforhealth.net/temp/EducHealth243493-8598473_235304.pdf)
15. Heijne-Penninga M, Kuks JBM, Hofman WHA, Muijtjens AMM, Cohen-Schotanus J. Influence of PBL with open-book test on knowledge retention measured with progress tests. *Advances in Health Sciences Education*. 2013; 18:485–495.
16. Johnson TR, Khalil MK, Peppler RD, Davey DD, Kibble JD. Use of the NBME comprehensive basic science examination as a progress test in the curriculum of a new medical school. *Advances Physiology Education*. 2014; 38:315–320.
17. Cecilio-Fernandes D, Kerdijk W, Jaarsma ADC, Tio RA. Development of cognitive processing and judgments of knowledge in medical students: analysis of progress test result. *Medical Teacher*. 2016; 38(11):1125–1129.
18. De Champlain AF, Cuddy MM, Scoles PV, Brown M, Swanson DB, Holtzman K, Butler A. Progress testing in clinical science education: results of a pilot project between the National Board of Medical Examiners and a US medical school. *Medical Teacher*. 2010; 32:503–508.
19. Williams RG, Klamen DL, White CB, Petrusa E, Fincher RME, Whitfield CF, Shatzer JH, McCarty T, Miller BM. Tracking development of clinical reasoning ability across five medical schools using a progress test. *Academic Medicine*. 2011; 86(9):1148–1154.
20. Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessment: feedback for individuals and institutions based on the progress test. *Medical Teacher*. 2010; 32(6):486–490.
21. Kerfoot BP, Shaffer K, McMahon GT, Baker H, Kirdar J, Kanter S, Corbett Jr EC, Berkow R, Krupat E, Armstrong EG. Online “Spaced Education Progress-Testing” of Students to Confront Two Upcoming Challenges to Medical School. *Academic Medicine*. 2011;86(3):300-306.
22. Wagener S, Möltner A, Timbil S, Gornostayeva M, Schultz JH, Brüstle P, Mohr D, Beken AV, Better J, Fries M, Gottschalk M, Günther J, Herrmann L, Kreisel C, Moczko T, Illg C, Jassowicz A, Müller A, Niesert M, Strübing F, Jünger J. Development of competency-based formative progress test with student-generated MSQs: Results from a multi centre pilot study. *GMS Zeitschrift für Medizinische Ausbildung*. 2015; 32(4):1–29.
23. van Diest R, van Dalen J, Bak M, Schruers K, van der Vleuten C, Muijtjens A, Scherpbier A. Growth of knowledge in psychiatry and behavioural sciences in a problem-based learning curriculum. *Medical Education*. 2004 Dec; 38(12):1295–1301.
24. Given K, Hannigan A, McGrath D. Red, yellow and green: what does it mean? How the progress test informs and supports student progress. *Medical Teacher*. 2016; 38(10):1025–1032.