Original Article

Correlation of the scores on different comprehensive examinations and the Medical Competency Assessment Test for National License: a mixed methods study

Weeratian Tawanwongsri, Tharin Phenwan

Abstract

Introduction: In-house comprehensive examinations for preclinical students from various Thai medical schools were provided. However, their quality has never been evaluated with a score correlation with NLE as the gold standard. This study aimed to assess the correlation of the scores as determined by three different comprehensive examinations and national license examination 1 (NLE step I) and to identify the positive learning strategies.

Method: A mixed methods sequential explanatory study was done to investigate the performances amongst four tests and to identify the potential factors affecting those scores using semi-structured interviews and focus groups with content analysis.

Result: All (n = 48) third-year medical students participated in our study. The majority were females (64.6%). Significantly positive correlations of NLE scores were: test A scores (r = 0.86), test B scores (r = 0.85), and WU test scores (r = 0.78). The highest accuracy index (AI = 0.87) was the WU test, where sensitivity, specificity were 20.0% and 97.1%, respectively. The WU test revealed that it was most helpful in preparing them for the NLE. Students who passed the exam used three study strategies; group study, tutorial sessions, and review by themselves.

Discussion and Conclusion: There were strong positive correlations between three different in-house developed comprehensive examinations and NLE. The WU test showed the highest accuracy index to predict the NLE result. Regular review of lessons was emphasized as a cornerstone.

Keywords: Comprehensive examination, Thai, National license examination, Correlation, Performance

Received: 26 June 2016 **Accepted: 9 November 2017**

School of Medicine, Walailak University

Corresponding author: Weeratian Tawanwongsri School of Medicine, Walailak University, Thaiburi, Thasala, Nakhon Si Thammarat, Thailand 80161 E-mail: weeratian.ta@gmail.com

Introduction

In the same way as getting medical licensure in the United States which consists of three steps designed to assess a physician's ability to apply a broad spectrum of knowledge to evaluate the physician's basic patient-centered skills, Thai medical students have to pass a three-step examination provided by the Center for Medical Competency Assessment and Accreditation (CMA) of Thailand[1]. Third-year medical students are eligible for step 1 NLE enrolment. It contains 300 multiple choice questions which are designed to assess how well the students apply basic scientific principles to relevant clinical cases. The time allowed for this examination is 6 hours.

Furthermore, another examination, called the comprehensive examination, was provided by some medical schools in order to judge whether students are eligible for studying in clinical years. Our school provides students with one in-house developed examination, and two other examinations obtained from two different well-known medical schools. The students have to pass at least one examination. However, the quality of these three comprehensive examinations has never been compared with one another or evaluated with a score correlation with NLE as the gold standard. Previous studies have reported a wide range correlation coefficient (r = 0.57 - 0.83) between the scores of in-house comprehensive examinations from various medical schools and those of The United States Medical Licensing Examination step 1 (USMLE step 1)[2 - 8]. Moreover, one study reported the in-house examinations were of relatively lower quality owing to the hugely different scores compared with the national examination[9].

Little is known about which study habits are associated with a higher score. Students who studied for 8 - 11 hours or more per day had higher scores, but there was no added benefit with additional study time. Those who completed an estimated > 2000 practice questions also obtained higher scores[10]. Even though previous research has documented that students who do aerobic exercise and have an adequate sleep consistent with Centers for Disease Control and Prevention guidelines appear less likely to experience burnout and to have a higher quality of life, health habits which are related to a high scores have never been well established[11, 12].

The aims of this study were 1) to assess the correlation of scores as determined by three different comprehensive examinations and NLE and 2) to identify the learning strategies and health habits students used to reach the Minimum Pass Level of NLE.

Method

To assess the correlation of scores on three different comprehensive examinations and the NLE as a national standard test and identify the potential factors which affect the NLE scores, a mixed method study was used. We use mixed methods sequential explanatory design to investigate students' performances among four mentioned tests and also their learning habits. All third-year medical students were included in this study. Participants' characteristics were obtained. Score data were collected from four examinations in the academic year 2016. The first two examinations were obtained from two different medical schools, i.e. test A and test B. The others were the in-house developed examinations, the WU test, and NLE provided by the CMA. To pass test A or test B or the WU test, 60% of the maximum score had to be attained. A 53.18% of NLE score was stipulated as the minimum passing level in this academic year. After students had finished all four tests, we used a purposive sampling method to select students who scored high in the tests with high GPAX, then conducted semi-structured interviews and focus group interviews to explore more about their learning strategies to prepare for the tests. Interviewing topics were; 1) their learning strategies and health habits, 2) their opinions regarding each comprehensive test.

All interviews were voice-recorded and transcribed verbatim. One researcher (TP) conducted all of the interviews. Data were gathered until we reached a saturation point where no new concepts emerged.

Examination organization

The examinations, except the WU test, were a six-hour assessment, and were separated into a morning and afternoon session. Each examination consisted of 300 English multiple-choice questions: 150 questions in the morning session and 150 questions in the afternoon session. As for the WU test, it consisted of 300 English multiple-choice questions, which were split into three parts, i.e. 100 questions each, on different days. Each part lasted 2 hours. They were not dependent on each other. Test A and Test B were held 7 weeks and 4 weeks prior to the NLE, respectively. WU test part 1, part 2, and part 3 were held 18 weeks, 13 weeks, and 2 weeks prior to the NLE, respectively. Similarly, each correct answer was equal to one mark in all tests.

Statistical analysis

The statistical analysis was performed using SPSS software version 17 (SPSS Inc., Chicago, IL, USA). Mean and standard deviation (SD) or median and range were used to describe continuous data. Frequency and percentage were used for categorical data. Analyses of data were performed using Mann-Whitney U test, student t-test, or one-way ANOVA test depending on data distribution. A p-value of < 0.05 by two-tailed tests was considered statistically significant. To assess the examination performances, linear regressions were applied to examine the quantitative correlations among the four examinations. For the qualitative analysis, we used methodological methods to increase the rigor of this work and used a content analysis approach to analyze the data. Codes were analyzed with Atlas.Ti 8.0 software.

Result

All third-year medical students participated in our study (n = 48). The students' characteristics were presented in Table 1. The students' mean age was 21.2 ± 0.5 years. Female students made up the majority of participants (n = 31, 64.6%). The mean Grade Point Average (GPAX) of male and female students was 3.42 (SD 0.30) and 3.43 (SD 0.27), respectively. There was no significant difference between male and female students with regard to GPAX. Forty-eight students (100%) enrolled for Test A and Test B. While, thirty-nine students (81.3%) enrolled in all three parts of the WU test.

Table 1 Student's characteristics

Characteristics	n = 48
Gender	
Male, n (%)	17 (35.4)
Female, n (%)	31 (64.6)
Age, mean (SD)	21.2 (0.5)
GPAX*	
Male, mean (SD)	3.42 (0.30)
Female, mean (SD)	3.43 (0.27)
Enrollment	
Test A, n (%)	48 (100)
Test B, n (%)	48 (100)
WU test, n (%)	39 (81.3)

*p-value = 0.90

All scores in the datasets followed a normal distribution determined by a Shapiro-Wilk test. Mean scores of test A, test B, WU test and NLE were 158.13 (SD 22.10), 148.48 (SD 20.12), 145.59 (SD 19.04), and 175.73 (SD 3.51), respectively.

Table 2 Students' scores on test A, test B, WU test and NLE.

Tests	Mean score	Standard deviation	MPL
Test A1	58.13	22.10	118.20
Test B1	48.48	20.12	122.40
WU test	145.59	19.04	113.40
NLE	175.73	3.51	159.01

MPL minimum pass level

Mean difference scores were calculated. Students' mean scores on NLE are 28.20, 37.85, and 40.26 significantly higher than the mean scores on test A, test B, and the WU test, respectively. Using the aforementioned passing criteria, each test performance in order to predict NLE fail status was given in Table 3 with sensitivity, specificity, and accuracy index.

Table 3 Sensitivity and specificity of tests in terms of determining students' NLE fail status.

	Sensitivity (%) (95%CI)	Specificity (%) (95%CI)	PPV (%) (95%CI)	NPV (%) (95%CI)	Accuracy index
Test A	100.00 (59.04 - 100.00)	41.46 (26.32 - 57.89)	22.58 (18.40 - 27.40)	100.00 (100.00 - 100.00)	0.50
Test B	85.71 (42.13 - 99.64)	75.61 (59.70 - 87.64)	37.50 (24.44 - 52.68)	96.87 (83.36 - 99.48)	0.77
WU test	20.00 (0.51 - 71.64)	97.06 (84.67 - 99.93)	50.00 (6.86 - 93.14)	89.19 (84.13 - 92.77)	0.87

According to the highest accuracy index, an ROC curve analysis of the WU test was further performed. The area under the curve (AUC) was 0.86, which demonstrated moderate discriminatory power, with SE = 0.08 and 95%CI from 0.70 to 1.00. Considering Fig. 1B, for students scoring above 173.5, the sensitivity and specificity of the WU test compared to the NLE was 91.2% and 100.0%, respectively. For students scoring above 171.5, the sensitivity and specificity of the WU test compared to NLE was 88.2% and 100.0%.

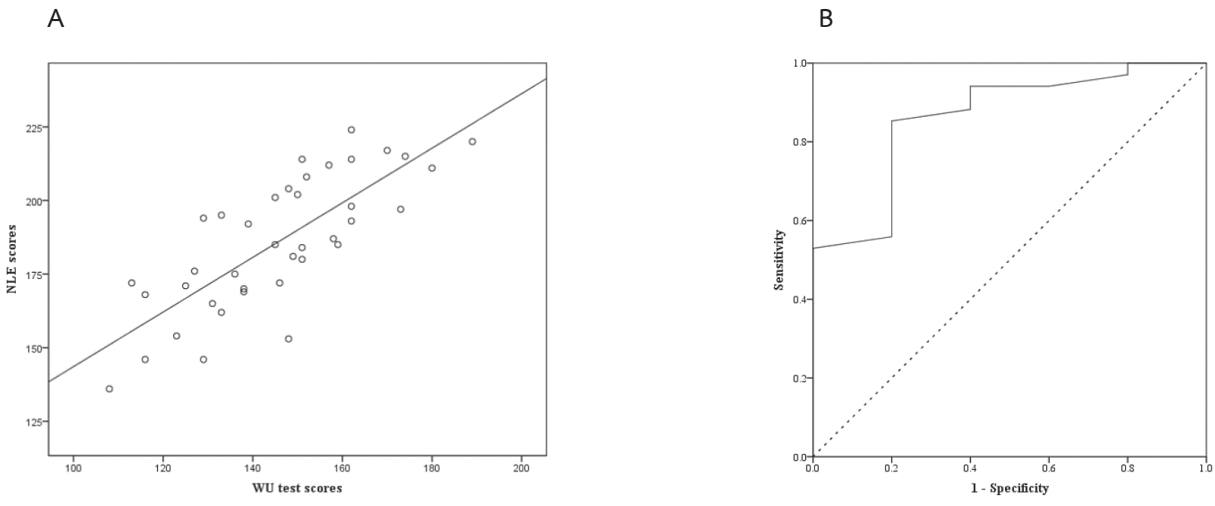

Figure 1 The linear regression analysis of students' scores on WU test and NLE (A) and the ROC curve using the WU test (B).

To predict more accurate NLE scores, a multiple linear regression analysis with categorical variables was done as seen in Table 4. Ultimately, the actual NLE scores are given by the following predictive model: 0.578 (Test A scores) + 38.088 (GPAX) - 35.784.

Table 4 Model used to predict accurate NLE score.

Potential factors	Unstandardized beta weights	Standard error	p-value
Test A scores	0.58	0.137	< 0.001
GPAX*	38.09	11.269	0.002

* Grade Point Average (GPAX)

Qualitative Results

Four students joined the semi-structured interviews and six students joined the focus groups (three students per session). The interviews lasted for 20 - 28 minutes and the focus groups lasted 39 and 58 minutes. According to learning strategies, all students (n = 10) used a mixture of study methods; group learning, tutorial sessions, and review by themselves. They stated that each method has its benefit and synergizes with each other. Firstly, students formed a study group at 10 - 12 months before the NLE, reviewed the topic individually, and shared with the study group.

"I think each method has its own good. For starters, it's good to start with group study." "Yeah. Because we can share on what we read with our classmates, saving times and energies." "But later on, I think reading by myself is better because we have to focus on what we find lacking which could be different topics."

Student 1 - 3/Focus group 1

"I started reading seriously after New Year but we already have study group during our second year. Around... last May or June."

Student 2/ Focus group 1

"I started [reading] when our seniors finished their [NLE] test. When I saw their results, I realised that it's our time now."

Student 7/Focus group 2

Three students found that in order to make this learning method effective, each member has to study on their parts thoroughly.

"[tutorial sessions] You have to be thoroughly prepared to make it most effective."

Student 9

Owning to health habits and extra-curricular activities, all students were actively involved in extra-curricular activities and exercised regularly.

"We play badminton 3 - 5 days a week." "[interviewer] Do you think these extra activities affect your test preparation?" "[looked confused] Not really." "No. I usually read at night time so I can manage my time." I play basketball regularly and agree with him. It's up to how we prioritize our time that matters."

Student 6 - 8/Focus group 2

For the extra activities, students who prepared for the test regularly-regular type reader-agreed that the activities did not affect their study as long as they managed their time properly.

[preparing for NLE] "I will go back to my room, exercise, go to the club, then come back to read. Make it a routine and a part of your day."

Student 10

As for the sleeping pattern, all of them slept for 6-8 hours.

"I try to sleep at least 6 - 7 hours every day or else I would feel really cranky in the morning and it affects my study."

Student 1

For Test A, all students (n = 10) reflected that the test was too complex, emphasizing basic science aspects and were not similar to the actual NLE in any way. They also had to take this test much earlier than the others while they had not finished the third-year curriculum yet.

"I don't think we would benefit from doing it [test A]. The test was different from what we studied and does not relate to the real test [NLE] at all." "I agree. Plus, we have to take the test when we have not finished our third year yet. Even if we did good or not, it does not reflect our capacities."

Student 1 and 2/Focus group 1

"Mainly memorization." "Yup. You don't have to analyze anything, if you can memorize it, you can do it. And it's not related to NLE at all."

Student 7 and 8/Focus group 2

Eight students found test B was more similar to NLE than test A, focusing on clinical correlations and what they studied.

"It's more case-based. Much like NLE."

Student 5

For our in-house test, i.e. the WU test, eight students gave positive feedback that the test helped them prepare for the NLE for two reasons. Firstly, it simulated test environment. Secondly, it helped them to prepare for the upcoming test.

"It helps immensely. The atmosphere is so realistic, with the stop watch and stuff. When I did the NLE, I wasn't very nervous. I think it's partly to the pretest."

Student 4

Discussion and Conclusion

Our first objective was to assess the correlation of scores on three different comprehensive examinations and the NLE as a national standard test. The correlation among these tests has never been declared in Thailand. Our paper presented a novel view for medical schools to pay attention to developing an in-house examination with the national standard in terms of determining whether their students are qualified. Previous studies in the United States of America revealed USMLE step 1 scores significantly correlated with the Comprehensive Basic Science Examination (CBSE) and the CBSE is a useful tool for the identification of students at risk of failing the USMLE Step 1[6, 13 - 17].

As was stated in the results, there were strongly positive correlations between three different comprehensive examination scores and NLE scores. The performance of independent WU test demonstrated the highest accuracy index of 0.87 with the highest specificity of 97.06%, coinciding with qualitative analysis of the overwhelming preference for the WU test over the other tests for its objective feedback. As a result of linear regression analysis, the line of regression of NLE scores on WU test scores was given by NLE score = 50.83 + 0.93 (WU test score) which was used to predict the NLE scores using the known WU test scores. Still, despite the strong positive correlation of each test with NLE, content analysis from students showed that test A might prove to be the least beneficial due to the nature of the test that focused on the memorization of basic science aspects. Students also had to take test A very early when they had not finished their third-year curriculum yet. Revision of test A or postponing of the test after they had finished the study was recommended.

For the learning strategies, apart from multiple study methods, all students emphasized the importance of preparing the test material regularly. They reflected that being a regular reader helped them prepare for the test better. All of them also exercised regularly and tried to sleep for at least 6 - 8 hours to maintain their quality of life and prevent unnecessary burnout, supporting previous studies[11, 12]. Based on these baseline characteristics, we also found no correlation between gender and the NLE pass/fail status. However, there was a modest correlation between the students' GPAX and the NLE pass/fail status (correlation coefficient 0.55, p-value < 0.001). The mean GPAX of the fail group and pass group was 3.05 (SD 0.17) and 3.48 (SD 0.24) with a significant difference in means of 0.50 (p-value 0.01). Nevertheless, further research will be required to investigate the difference in learning strategies between the pass group and the fail group in order to implement strategies to enhance student learning, particularly the fail group.

To our knowledge, this was the first study that analyzes the correlation of each in-house test with the standard test, i.e. NLE. We also used a mixed-methods approach to delve deeper into learning strategies and health habits of the students to strengthen the quantitative results. Results would be used to improve the test in upcoming future. Our research had two limitations. The first is missing data of the WU test. There were 39 students (81.3%) who enrolled and took this examination. Our collected data revealed that there were two students who failed the NLE and seven students who passed the NLE. It is plausible that this could have influenced and made the correlation results incomplete or inaccurate. The second is the timing of the examinations as detailed in the examination organization. Students' preparedness and readiness might vary at different times. We assume that the ideal time in order to evaluate these correlations is when the NLE is held. Practically, we are not able to arrange these four examinations on the same day due to student fatigue. On the contrary, one previous study revealed that the number of days studied did not have any correlation with the scores, suggesting that increased length of study may not ameliorate poor grades[13].

In conclusion, our work revealed strongly positive correlations between three different in-house developed comprehensive examinations and NLE with our independent test showed the highest accuracy index to predict the NLE result with a high correlation. There was a modest correlation between students' GPAX and the NLE pass/fail status. All students who passed the exam used multiple learning strategies, reviewing regularly and exhibit healthy exercise habits. Further research to maximize students' learning strategies especially in those who fail the test is recommended.

Funding

Institutional Research Section, Division of Planning and Strategy, Walailak University.

Competing interests

The authors declare that they have no competing interests.

Ethics approval

The Walailak ethical committee of the institute has approved the study protocol (WUEC-16-024-01). The study complied with the International Conference on Harmonization of Good Clinical Practice and principles of the Declaration of Helsinki.

References

1. Comprehensive Basic Medical Sciences examination contents Bangkok: Center for Medical Competency Assessment and Accreditation; [updated 28 NOV 2014. Available from: http://www.cmathai.org/news/practical/243.
2. Coumarbatch J, Robinson L, Thomas R, Bridge PD. Strategies for identifying students at risk for USMLE step 1 failure. Fam Med. 2010;42(2):105-10.
3. Davis GE, Gayer GG. Comparison of Basic Science Knowledge Between DO and MD Students. J Am Osteopath Assoc. 2017;117(2):114-23.
4. Denton GD, Durning SJ, Wimmer AP, Pangaro LN, Hemmer PA. Is a faculty developed pretest equivalent to pre-third year GPA or USMLE step 1 as a predictor of third-year internal medicine clerkship outcomes? Teach Learn Med. 2004;16(4):329-32.
5. Gauer JL, Wolff JM, Jackson JB. Do MCAT scores predict USMLE scores? An analysis on 5 years of medical student data. Med Educ Online. 2016;21(1):31795.
6. Giordano C, Hutchinson D, Peppler R. A Predictive Model for USMLE Step 1 Scores. Cureus. 2016;8(9):e769.
7. Sarko J, Svoren E, Katz E. COMLEX-1 and USMLE-1 are not interchangeable examinations. Acad Emerg Med. 2010;17(2):218-20.
8. Slocum PC, Louder JS. How to predict USMLE scores from COMLEX-USA scores: a guide for directors of ACGME-accredited residency programs. J Am Osteopath Assoc. 2006;106(9):568-9.
9. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med. 2002;77(2):156-61.
10. Kumar AD, Shah MK, Maley JH, Evron J, Gyftopoulos A, Miller C. Preparing to take the USMLE Step 1: a survey on medical students' self-reported study habits. Postgrad Med J. 2015;91(1075):257-61.
11. Dyrbye LN, Satele D, Shanafelt TD. Healthy Exercise Habits Are Associated With Lower Risk of Burnout and Higher Quality of Life Among U.S. Medical Students. Acad Med. 2016.
12. Wolf MR, Rosenstock JB. Inadequate Sleep and Exercise Associated with Burnout and Depression Among Medical Students. Acad Psychiatry. 2016.
13. Glew RH, Ripkey DR, Swanson DB. Relationship between students' performances on the NBME Comprehensive Basic Science Examination and the USMLE Step 1: a longitudinal investigation at one school. Acad Med. 1997;72(12):1097-102.
14. Morrison CA, Ross LP, Fogle T, Butler A, Miller J, Dillon GF. Relationship between performance on the NBME Comprehensive Basic Sciences Self-Assessment and USMLE Step 1 for U.S. and Canadian medical school students. Acad Med. 2010;85(10 Suppl):S98-101.
15. Burns ER, Garrett J. Student failures on first-year medical basic science courses and the USMLE step 1: a retrospective study over a 20-year period. Anat Sci Educ. 2015;8(2):120-5.

16. Sawhill A, Butler A, Ripkey D, Swanson DB, Subhiyah R, Thelman J, et al. Using the NBME self-assessments to project performance on USMLE Step 1 and Step 2: impact of test administration conditions. Acad Med. 2004;79(10 Suppl):S55-7.

17. Swanson DB, Ripkey DR, Case SM. Relationship between achievement in basic science course-work and performance on 1994 USMLE Step 1. 1994-95 Validity Study Group for USMLE Step 1/2 Pass/Fail Standards. Acad Med. 1996;71(1 Suppl):S28-30.

บทคัดย่อ

ความสอดคล้องของคะแนนการสอบประมวลความรู้ประเภทต่างๆ และคะแนนการสอบเพื่อประเมินและรับรองความรู้ความสามารถในการประกอบวิชาชีพเวชกรรม: การวิจัยแบบผสม

วีรเธียร ถวัลย์วงศ์ศรี, ธารินทร์ เพ็ญวรรณ

สำนักวิชาแพทยศาสตร์ มหาวิทยาลัยวลัยลักษณ์

ผู้ให้ติดต่อ: อาจารย์นายแพทย์วีรเธียร ถวัลย์วงศ์ศรี สำนักวิชาแพทยศาสตร์ มหาวิทยาลัยวลัยลักษณ์ ๒๒๒ ตำบลไทยบุรี อำเภอท่าศาลา จังหวัดนครศรีธรรมราช รหัสไปรษณีย์ ๘๐๑๖๐ อีเมล: weeratian.ta@gmail.com

บทนำ: เพื่อประเมินความสอดคล้องของคะแนนสอบข้อสอบประมวลความรู้จากโรงเรียนแพทย์สามแห่ง เปรียบเทียบกับคะแนนสอบข้อสอบขั้นตอนที่ ๑ ของศูนย์ประเมินและรับรองความรู้ความสามารถในการประกอบวิชาชีพเวชกรรม (ศ.ร.ว.) และระบุวิธีการเตรียมตัวสอบของนักศึกษาแพทย์ที่สอบผ่าน

วิธีการศึกษา: เป็นการวิจัยแบบผสม ระบุระดับความสัมพันธ์จากค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนสอบของนักศึกษาแพทย์ (นศพ.) จากชุดข้อสอบทั้งสี่ฉบับ และใช้การสัมภาษณ์เพื่อระบุวิธีการเตรียมตัวสอบของ นศพ. ที่สอบผ่าน

ผลการศึกษา: นศพ. ชั้นปีที่สาม ๔๘ คน ซึ่งส่วนใหญ่เป็นเพศหญิง (ร้อยละ ๖๔.๖) การวิเคราะห์สหสัมพันธ์ พบว่าคะแนนสอบข้อสอบประมวลความรู้และคะแนนสอบของ ศ.ร.ว. มีความสัมพันธ์แบบตามกันอย่างมีนัยสำคัญทางสถิติ (โรงเรียนแพทย์ ก, r = ๐.๘๖; โรงเรียนแพทย์ ข, r = ๐.๘; และมหาวิทยาลัยวลัยลักษณ์ (มวล.), r = ๐.๗๘) ข้อสอบ มวล. มีความแม่นยำสูงสุดเพื่อพยากรณ์การสอบตกข้อสอบของ ศ.ร.ว. (AI = ๐.๘๗) มีความไวและความจำเพาะร้อยละ ๒๐.๐ และร้อยละ ๙๗.๑ ตามลำดับ นักศึกษาแพทย์ที่สอบผ่านเตรียมตัวสอบด้วยการทบทวนแบบกลุ่ม การสอนเพื่อน และทบทวนด้วยตนเอง

วิจารณ์ และสรุปผลการศึกษา: คะแนนสอบข้อสอบประมวลความรู้ทั้งสามฉบับและคะแนนสอบข้อสอบ ศ.ร.ว. มีความสัมพันธ์แบบตามกันอย่างมีนัยสำคัญทางสถิติ ข้อสอบ มวล. มีความแม่นยำสูงสุดเพื่อพยากรณ์การสอบตกข้อสอบของ ศ.ร.ว. โดยเตรียมตัวสอบอย่างสม่ำเสมอเป็นสิ่งสำคัญ

คำสำคัญ: การสอบประมวลความรู้, ไทย, การสอบประเมินและรับรองความรู้ความสามารถในการประกอบวิชาชีพเวชกรรม, ความสัมพันธ์