Chula Med J Vol. 46 No. 10 October 2002

A comparative study of PS-2 and SB-8 grading programs employed at the Faculty of Medicine, Chulalongkorn University

Boonnart Laisnitsarekul*

Laisnitsarekul B. A comparative study of PS-2 and SB-8 grading programs employed at the Faculty of Medicine, Chulalongkorn University. Chula Med J 2002 Oct; 46(10): 811 - 22

Objective : *To compare grading results of medical students' examinations, Faculty of Medicine, Chulalongkorn University, derived from two grading programs: PS-2 and SB-8, in terms of number of student in each grade (A, B^+, B, C^+, C, D^+, D, F) and time spent in their calculations.*

Design : *Experimental study.*

Methods : *The researcher used MCQ scores of Comprehensive Examination for sixth year medical students, Faculty of Medicine, Chulalongkorn University, academic years 1993 to 2000. Scores were graded by PS-2 and SB-8 grading program under the same condition. Time spent in their calculations were also counted and tested difference for the number of student in each grade by Chi-square test.*

Results : *The eight Comprehensive MCQ tests, academic years 1993 - 2000, had 298 to 300 total scores. The number of students were between 136 to 196. The reliability of the tests were 0.84 -0.87. The number of students in each grade calculated by PS-2 and SB-8 from 1993-2000 Comprehensive MCQ test were not significantly difference. When using PS-2 included Standard Error of Mean and SB-8 included Standard Error of Measurement, there was significantly difference. This research showed 2 essential results. The first result is the SB-8 grading program is a good program as same as the PS-2*

*Medical Education Unit, Faculty of Medicine, Chulalongkorn University

grading program, the Faculty standard. The second result is the weak point of the PS-2 in terms of using mean and S_M in program. It should be better and appropriate if the PS-2 uses median and SEM.

Summary : *The experimental research is aimed to compare the number of students in each grade (A, B^+, B, C^+, C, D^+, D, F) given by the two grading programs, PS-2 and SB-8. The scores of 8 Comprehensive MCQ examinations (1993 - 2000), Faculty of Medicine, Chulalongkorn University were employed as their input database. The reliability of the Comprehensive MCQ tests were 0.84-0.87. The number of the sixth year medical students were 136 -196. The number of students in each grade were not significantly different except when compared PS-2 included Standard Error of Mean and SB-8 included Standard Error of Measurement.*

Key words : *Grading, PS-2, SB-8, MCQ, Comprehensive Examination, Standard Error of Measurement (SEM), Standard Error of Mean (S_m).*

Reprint request : Laisnitsarekul B. Medical Education Unit, Faculty of Medicine, Chulalongkorn University, Bangkok 10330,Thailand.

Received for publication. May 5,2002.

บุญนาท ลายสนิทเสรีกุล. การเปรียบเทียบโปรแกรมตัดเกรด PS-2 และ SB-8 ของคณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. จุฬาลงกรณ์เวชสาร 2545 ต.ค ; 46(10): 811 - 22

วัตถุประสงค์ : *เพื่อเปรียบเทียบผลการตัดเกรดนิสิตแพทย์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ระหว่างโปรแกรม PS-2 กับโปรแกรม SB-8 ในแง่จำนวนนิสิตที่ได้แต่ละเกรด (A, B^+, B, C^+, C, D^+, D, F) และระยะเวลาที่ใช้ในการคำนวณ*

รูปแบบการวิจัย : *การวิจัยเชิงทดลอง*

วิธีการศึกษา : *ผู้วิจัยใช้ผลการสอบวิชาเวชปฏิบัติทั่วไปชนิดข้อสอบปรนัย ของนิสิตแพทย์ชั้นปีที่ 6 ปีพุทธศักราช 2536-2543 นำมาตัดเกรดด้วยโปรแกรม PS-2 และ SB-8 โดยใช้เงื่อนไขเดียวกัน*

ผลการศึกษา : *ผลการศึกษาพบว่า ข้อสอบวิชาเวชปฏิบัติทั่วไปชนิดข้อสอบปรนัย ปีการศึกษา 2536-2543 มีค่าความเที่ยงระหว่าง 0.84-0.87 จำนวนนิสิตแพทย์ชั้นปีที่ 6 อยู่ระหว่าง 136-196 คน เมื่อนำคะแนนสอบมาตัดเกรดด้วยโปรแกรม PS-2 และ SB-8 จำนวนนิสิตแพทย์ในแต่ละเกรดไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ถ้ามีการนำค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) มาคำนวณด้วย จำนวนนิสิตแพทย์ในแต่ละเกรดที่คำนวณด้วยโปรแกรม SB-8 จะมีมากกว่าจำนวนนิสิตแพทย์ในแต่ละเกรดที่คำนวณด้วยโปรแกรม PS-2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($p < .05$) การวิจัยครั้งนี้มีข้อค้นพบสำคัญ 2 ประการ คือ โปรแกรม SB-8 มีคุณภาพดีเทียบเท่าโปรแกรม PS-2 ซึ่งเป็นโปรแกรมมาตรฐานของคณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และจุดอ่อนของโปรแกรม PS-2 ที่ใช้ค่าสถิติมัชฌิมเลขคณิต และค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย ซึ่งหากใช้ค่ามัธยฐาน และค่าความคลาดเคลื่อนมาตรฐานในการวัด จะมีความถูกต้องสมบูรณ์ยิ่งขึ้น*

สรุป : *งานวิจัยเชิงทดลองนี้มีวัตถุประสงค์เพื่อเปรียบเทียบจำนวนนิสิตแพทย์ในแต่ละเกรด (A, B^+, B, C^+, C, D^+, D, F) เมื่อใช้โปรแกรมการตัดเกรดระหว่างโปรแกรม PS-2 และโปรแกรม SB-8 คะแนนที่นำมาตัดเกรดเป็นคะแนนข้อสอบปรนัยของวิชาเวชปฏิบัติทั่วไป ปีพุทธศักราช 2536 - 2543 คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย รวม 8 ชุด ความเที่ยงของข้อสอบปรนัยทั้ง 8 ชุด อยู่ระหว่าง 0.84-0.87 จำนวนนิสิตแพทย์ชั้นปีที่ 6 อยู่ระหว่าง 136-196 คน ผลการเปรียบเทียบพบว่าจำนวนนิสิตแพทย์ในแต่ละเกรด เมื่อใช้โปรแกรมการตัดเกรด PS-2 และ SB-8 ไม่แตกต่างกัน ยกเว้นเมื่อใช้ค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย (Standard Error of Mean) มาคำนวณด้วยในโปรแกรม PS-2 เปรียบเทียบกับโปรแกรม SB-8 ที่นำค่าความคลาดเคลื่อนมาตรฐานในการวัดมาคำนวณด้วย จะแตกต่างกันอย่างมีนัยสำคัญทางสถิติ*

Concerning its grading systems, the letter marks are usually used in academic institutions. The Faculty of Medicine, Chulalongkorn University used five categories (i.e. A, B, C, D, F) since 1974 [1] and later switched to eight categories (i.e. A, B^+, B, C^+, C, D^+, D, F) in 1997.[2] Arunyingmongkol and Laisnitsarekul constructed a computer program for the Eight-Category Grading System named SB-8 in 1997.[3] In 1998 Siriviriyakul and Kamolwarin constructed a computer program for the Eight-Category Grading System named PS-2 in the name of the Division of Academic Affairs, Faculty of Medicine, Chulalongkorn University.[4] The strong points of the PS-2 grading program were 1) easy to understand because the program was written in Thai language and 2) familiar for user because it constructed from the Microsoft Excel. However the weak points of the PS-2 were 1) used arithmetic mean as average 2) the interval between each grade depended on user. The strong points of the SB-8 grading program were 1) available to use with many computers such as PC386, PC486, Pentium-I, Pentium-II, Pentium-III, Pentium-4 2) used median as average 3) the interval between each grade calculated by computer and 4) there were standard error of measurement in the SB-8. The weak point was English language of the program. It might be not familiar for user. This stimulated the author's interest in studying the difference in number of student in each grade between the two aforementioned grading systems.

Objectives

1. To calculate the grade from the scores of tests by the PS-2 and SB-8 grading programs.

2. To find and compare the time spent in preparing and processing the data.

3. To compare the difference in number of students in each grade between the abovementioned grading programs.

Hypothesis

No difference in number of student in each grade between the PS-2 and SB-8 grading system programs.

Definitions

1. Standard Error of Measurement (SEM): The standard error of measurement provides an estimate of the range of variation in a set of repeated measurements of the same thing.[5] If it were possible to test a pupil over and over again on the same test, we would find that his scores would slightly vary. The amount of the variation in his test scores would be directly related to the reliability of the testing procedures. Low reliability would be indicated by large variations in the pupil's test scores. High reliability would be indicated by little variation from one testing to another. Although it is impractical to administer a test many times to the same pupils, it is possible to estimate the amount of variation to be expected in test scores. This estimate is called the standard error of measurement.[6] To calculate, use the following formula:[7]

$$SEM = SD\sqrt{1-r}$$

Where SEM = Standard Error of Measurement

SD = Standard deviation of norm group

r = Reliability of test

2. Standard Error of the Mean (S_M): Since the extent and the distribution of sampling errors can be predicted, we can use sample means with predictable confidence to make inferences concerning population means. However, we need an estimate of the magnitude of the sampling error associated with the sample mean when it is used as an estimate of the population mean. An important tool for this purpose is the standard error of the mean.[8] The standard error of the mean is the standard deviation of the sampling distribution of the mean. The formula for the standard error of the mean is [9] :

$$S_M = \frac{S}{\sqrt{N}}$$

Where S_M = Standard Error of the Mean

S = Standard deviation of the original distribution

N = Sample size (the number of scores each mean is based upon)

Materials

1. One microcomputer Pentium CELERON 300 MHz.
2. One NEC-P6300 Dot Matrix 24 pins printer.
3. One OPSCAN Model 5 optical reader.
4. One stop-watch.
5. TOOLS : Software for the optical reader.
6. Two Grading system software programs : PS-2 and SB-8.
7. Word processing software program (QEdit).
8. Spread sheet software (Microsoft Excel).
9. Statistical software (EPISTAT).
10. Eight of the Comprehensive MCQ tests in academic year 1993-2000.

Methods

1. The optical reader scanned the students' answer sheets to obtain raw score. After scanning the students' answer sheets, the optical reader was given raw score as shown in Fig.1

ID	300
3346701	214
3346699	164
3346682	201
3346674	191
3346658	184
3346641	204
3346633	201
3346625	191
3346618	165
3346600	188
3346598	192
3346573	186
3346565	197

Figure 1. A sample of raw score created by OPSCAN.

2. Use the word processing software to prepare the raw score. The raw score was prepared as a data file for the two grading system programs. For the PS-2 and SB-8 programs, the raw score was prepared as SCORE4PS. DAT and SCORE4SB.DAT, respectively. A part of a SCORE4PS.DAT file was shown in Fig.2 and a part of a SCORE4SB.DAT file was shown in Fig.3

214
164
201
191
184
204
201
191
165
188

Figure 2. Format of data structure in a SCORE4PS.DAT file.

```
300
214
164
201
191
184
204
201
191
165
188
192
186
197
```

Figure 3. Format of data structure in a SCORE4SB.DAT file.

3. Run each grading system program and checked the time spent in 4 categories such as: 1) time spent in transferring data from OPSCAN to grading system program; 2) time spent in keying in data manually, 3) time spent in processing the database; and 4) time spent in printing the result.

4. Counted the numbers of grade A, B^+, B, C^+, C, D^+, D, F, and then compared the outcomes of two programs by Chi-square test.

Results

1. The eight Comprehensive MCQ tests in academic year 1993-2000 had 298 to 300 total scores. The number of students were 136 to 196. The reliability of tests were 0.84-0.87. The essential statistics were shown in Table 1.

2. The time spent in PS-2 grading system when analyzed the eight Comprehensive MCQ tests were: 23.45, 24.53, 21.35, 25.15. 26.40, 27.27, 29.12 and 27.37 minutes, respectively; their average was 25.58. The time spent for SB-8 grading system when analyzed the eight Comprehensive MCQ tests were: 10.34, 11.03, 8.18, 10.59, 10.50, 10.49, 12.56 and 11.02 minutes, respectively; their average was 10.59.

Table 1. The essential statistics of the Comprehensive MCQ test in academic year 1993 - 2000.

Item \ Academic	1993	1994	1995	1996	1997	1998	1999	2000
Number of students	148	154	136	160	174	178	196	189
Total scores	298	300	300	300	300	300	300	300
Maximum score	227	240	231	230	215	227	223	232
Minimum score	141	146	139	142	122	119	133	125
Mean	190.96	198.15	185.62	185.19	172.70	185.74	183.60	185.17
Median	193	198	186	187.50	172	186	184	186
Standard Deviation	17.55	17.23	16.90	18.94	19.70	18.95	18.05	17.57
Reliability (KR-20)	0.85	0.84	0.84	0.87	0.87	0.87	0.86	0.84
Standard Error of Mean	8.24	8.16	8.37	8.36	8.50	8.35	8.39	8.37
Standard Error of Measurement	6.90	6.89	6.83	6.99	7.19	6.86	6.89	7.05

Table 2. Time used in 4 categories between PS-2 and SB-8 for the 1993 -1996 Comprehensive MCQ test.

Time Used	1993		1994		1995		1996	
	PS-2	SB-8	PS-2	SB-8	PS-2	SB-8	PS-2	SB-8
Transfer data	1.20	0.51	1.30	0.50	1.10	0.24	1.40	0.40
Key data	5.43	5.43	6.08	6.08	4.25	4.25	6.0	6.0
Process data	1.57	0.30	2.15	0.35	1.50	0.24	1.55	0.29
Print result	14.45	3.30	15.00	3.30	14.10	3.05	15.40	3.50
Total	23.45	10.34	24.53	11.03	21.35	8.18	25.15	10.59

Table 2.(cont.) Time used in 4 categories between PS-2 and SB-8 for the 1997-2000 Comprehensive MCQ test.

Time Used	1997		1998		1999		2000	
	PS-2	SB-8	PS-2	SB-8	PS-2	SB-8	PS-2	SB-8
Transfer data	1.15	0.50	1.37	0.48	1.10	0.55	1.03	0.45
Key data	6.0	6.0	6.0	6.0	7.38	7.38	6.0	6.0
Process data	2.20	0.35	2.25	0.37	2.34	0.38	2.18	0.35
Print result	17.05	3.25	17.25	3.24	17.50	3.45	18.16	3.42
Total	26.40	10.50	27.27	10.49	29.12	12.56	27.37	11.02

Table 3.1 Number of students in each grade calculated by PS-2 and SB-8 and their significant difference from the 1993 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	28	31	21	47			
B^+	35	39	41	42			
B	36	33	37	29			
C^+	26	24	28	21	NS	NS	NS
C	15	13	13	8			
D^+	7	7	7	1			
D	1	1	1	-			
F	-	-	-	-			

Remark S_M = Standard Error of Mean
SEM = Standard Error of Measurement
NS = Not Significant

3. The number of students in each grade calculated by PS-2 and SB-8 from the 1993-2000, Comprehensive MCQ tests were not significantly difference. Upon applying PS-2 including Standard Error of Mean and SB-8, including Standard Error of Measurement, there were significantly difference at $p < .05$ to $p < .01$, except in the year 1993 Comprehensive MCQ test.

Table 3.2 Number of students in each grade calculated by PS-2 and SB-8 and their significant difference from the 1994 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	30	33	30	52			
B^+	30	30	39	48			
B	42	47	42	32			
C^+	30	23	30	14	NS	NS	P < .01
C	14	13	8	7			
D^+	7	7	4	1			
D	-	1	1	-			
F	1	-	-	-			

Table 3.3 Number of students in each grade calculated by PS-2 and SB-8 and their significant difference from the 1995 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	26	27	25	47			
B^+	36	38	40	40			
B	29	27	31	29			
C^+	28	27	27	13	NS	NS	P < .05
C	10	10	8	5			
D^+	4	5	4	2			
D	3	2	1	-			
F	-	-	-	-			

Table 3.4 Number of students in each grade calculated by PS-2 and SB-8 and their significant difference from the 1996 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	29	35	26	40			
B^+	41	40	38	56			
B	37	32	43	33			
C^+	26	30	30	16	NS	NS	P<.05
C	17	16	17	15			
D^+	10	7	6	-			
D	-	-	-	-			
F	-	-	-	-			

Table 3.5 Number of students in each grade calculated by PS-2 and SB-8 and their significant difference from the 1997 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	31	37	34	58			
B^+	40	34	40	50			
B	47	48	47	40			
C^+	30	34	38	19	NS	NS	P < .01
C	18	14	12	5			
D^+	6	5	3	2			
D	2	2	-	-			
F	-	-	-	-			

Table 3.6 Number of students in each grade calculated by the PS-2 and SB-8 and their significant difference from the 1998 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	31	34	30	58			
B^+	42	43	47	49			
B	39	47	51	45			
C^+	46	37	38	18	NS	NS	P < .01
C	12	11	7	4			
D^+	4	2	2	3			
D	2	3	2	1			
F	2	1	1	-			

Table 3.7 Number of students in each grade calculated by the PS-2 and SB-8 and their significant difference from the 1999 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	38	41	37	69			
B^+	42	47	46	55			
B	48	47	57	39			
C^+	35	28	33	27	NS	NS	P < .01
C	27	27	19	4			
D^+	4	4	3	2			
D	2	2	1	-			
F	-	-	-	-			

Table 3.8 Number of students in each grade calculated by the PS-2 and SB-8 and their significant difference from the 2000 Comprehensive MCQ test.

Grade	PS-2	PS-2+S_M	SB-8	SB-8+SEM	PS-2 vs SB-8	PS-2+S_M vs SB-8	PS-2+S_M vs SB-8+SEM
A	33	38	32	65			
B^+	46	46	47	52			
B	45	51	56	45			
C^+	38	30	32	17	NS	NS	P < .05
C	17	15	14	6			
D^+	6	5	6	3			
D	2	2	1	1			
F	2	2	1	-			

Discussion

Since the Comprehensive MCQ tests in academic year 1993 -2000 had high reliability, then the MCQ scores should be highly acceptable.[10-13] When using the grading programs named PS-2 and SB-8, the total time spent for PS-2 was longer than the total time spent for SB-8. The number of students in each categories (A, B^+, B, C^+, C, D^+, D, F) calculated by PS-2 and SB-8 were not significantly difference, except the number of students which calculated by PS-2 including Standard Error of Mean and SB-8 included Standard Error of Measurement. Based on the definition, it was questionable whether the Standard Error of Mean was suitable for grading system or not.

The purpose of evaluation is primarily the grading and classifying of students. It is designed to find those who have failed (D or F), those who have succeeded (A or B), and those who have gotten by (C).[14] The SEM for a test gives us and estimate of the amount of error (in test score units) associated with the score and individual receives on the test. The outstanding advantage of using the SEM is that it permits us to make certain probability statements about an individual's "true score" (the score he would obtain if there were no inconsistencies in the measuring system), given the score he actually received.[15] The SEM is a standard error of estimate and may be interpreted the limits marked off at distances of plus and minus 1 SEM from the regression line. In a test with a SEM equal to 2.0 units and if a certain individual's true score were 35, for example, the odds are 2 to 1 that his obtained score would not exceed 37 or fall below 33. Allowing a margin of 2 SEM, we can say that the odds are 19 to 1 that his obtained score will not exceed 39 or fall below 31.[16] When an instrument is applied repeatedly, we expect to observe some variation in the measurements recorded. The finer or smaller the average magnitude of this SEM, the more reliable the instrument.[17] Then There is SEM in the SB-8 grading system.

Young and Veldman[14] suggested that the mean should be used only when the distribution of scores was symmetrical. However, when the distribution was symmetrical, both the median and mean would be approximately the same. When the distribution was not symmetrical it was called a skewed distribution, the median was the proper measure of central tendency. Guilford and Fruchter[19] mentioned that if we are to use a sample mean as an estimate of the population mean, any deviation of such a sample mean from the population mean may be regarded as an error of estimation. The S_M tells us how large these error of estimation are in any particular sampling situation. The S_M is a standard deviation of the distribution of sample means.

This research showed 2 essential results. The first result is the SB-8 grading program is a good program as same as the PS-2 grading program, the Faculty standard. The second result is the weak point of the PS-2 in terms of using mean and S_M in program. It should be better and appropriate if the PS-2 uses median and SEM.

Summary

This experimental research was aimed to compare the outcomes of the number of students in each grade (A, B^+, B, C^+, C, D^+, D, F) by applying the two grading programs PS-2 and SB-8. The Eight Comprehensive MCQ scores (1993 - 2000), Faculty of Medicine, Chulalongkorn University were used. The reliability of the eight Comprehensive MCQ tests were: 0.84 - 0.87. The numbers of the sixth year medical students were: 136 -196. The number of students in each grade were not significant difference except when compared PS-2 including Standard Error of Mean, and SB-8 including Standard Error of Measurement.

References

1. Chulalongkorn University. The 1974 Chulalongkorn University Regulation for Educational System for Bachelors Degree. Bangkok : Division of Academic Affairs, 1974 : 7
2. Chulalongkorn University. The 1997 Chulalongkorn

University Regulation for Educational System for Bachelors Degree. Bangkok : Division of Academic Affairs, 1997 : 7

3. Arunyingmongkol S, Laisnitsarekul B. SB-8 : The eight-category grading program. Chula Med J 1998 Jul; 42 (7) : 587 - 97

4. Siriviriyakul P, Kamolwarin N. PS-2 : The eight-category grading program. Bangkok : Medical Education Unit, Division of Academic Affairs, 1999

5. Ary D, Jacobs LC, Razavieh A. Introduction to Research in Education. 2nd ed. New York : Holt Rinehart and Winston, 1979 : 217

6. Gronlund NE. Measurement and Evaluation in Teaching. 2nd ed. New York : Macmillan, 1971: 109

7. ASE realising potential. Standard error of measurement. [online] 2002. Available from: URL: http://www. ase-solutions.co.uk/html/advice/sem.htm

8. Ary D, Jacobs LC, Razavieh A. Introduction to research in education. 2nd.ed. New York : Holt, Rinehart and Winston 1979 : 138

9. Standard error of mean. In: Rice University. Hyperstat Online Textbook. [online] 2002. Available from:URL: http://davidmlane.com/hyperstat/A103735.html

10. Hubbard JP, Clemans WV. Multiple Choice Examinations in Medicine: A Guide for Examiner and Examinee. Philadelphia: Lea& Febiger, 1961: 71

11. Schumacher CF. Scoring and analysis. In : Hubbard JP, ed. Measuring Medical Education. Philadelphia : Lea & Febiger, 1971, 60 -1

12. Cox K, Ewan CE. The Medical Teacher. 2nd ed. London : Churchill Livingstone, 1988: 163

13. Gronlund NE. Measurement and Evaluation in Teaching. 2nd.ed. New York : Macmillan, 1971: 109

14. Bloom BS, Hastings JT, Madaus GF. Handbook on Formative and Summative Evaluation of Student Learning. New York : McGraw-Hill Book Company, 1971: 7

15. Whitla DK. Handbook of Measurement and Assessment in Behavioral Sciences. Massachusetts : Addison-Wesley Publishing Company, 1968 : 274

16. Guilford JP, Fruchter B. Fundamental Statistics in Psychology and Education. 5th ed. Tokyo: McGraw-Hill, Kogakusha, 1973: 402

17. Osgood CE, Suci GJ, Tannenbaum PH. The Measurement of Meaning. 8th ed. Chicago: University of Illinois Press, 1971: 129

18. Young RK, Veldman DJ. Introductory Statistics for the Behavioral Sciences. 2nd ed. New York: Holt, Rinehart and Winston, 1972: 63 - 7

19. Guilford JP, Fruchter B. Fundamental Statistics in Psychology and Education. 5th ed. Tokyo: McGraw-Hill, Kogakusha, 1973: 127