12-12-2014

The Test-Retest Reliability And Minimal Detectable Change In The Modified Fresno Test In Doctor Of Physical Therapy Students

Erin Pike
University of New England

Lisa Gerhardt
University of New England

The Test-Retest Reliability and Minimal Detectable Change in the Modified Fresno Test in Doctor of Physical Therapy Students

Erin Pike, BS and Lisa Gerhardt, BA

Students in the Doctor of Physical Therapy Program

University of New England, Portland, Maine

Approved by UNE Institutional Review Board

April 29, 2014

Acknowledgements:

Thank you to our faculty advisor Michael Fillyaw, PT, MS for guidance and participation completing this research study.

Abstract

Background: The American Physical Therapy Association identified the need for training in evidence based practice (EBP) and set forth guidelines for doctor of physical therapy (DPT) curricula to educate practitioners who are efficient and critical users of best evidence. Since DPT programs are teaching EBP, educators need an assessment tool to evaluate the competence of students. The Modified Fresno Test (MFT) of EBP was validated for physical therapists and the test-retest reliability and minimal detectable change (MDC) has been found for first year DPT students.

Objective: The purpose is to determine the test-retest reliability and MDC of the MFT in first, second, and third year DPT students. A secondary purpose is to compare the mean total score of the MFT among the three student groups.

Design: Test-retest design

Methods: Using a simple random sample, we recruited 21 University of New England (UNE) DPT students from each of the three classes. The participants completed the MFT twice, separated by 14 days, in a classroom on UNE's campus.

Results: Students in the third year class completed the validated 13-item MFT and due to a photocopying error, students in the first and second year class completed an 11-item MFT. The first year students had the lowest 11-item MFT mean score (68.5 points) which was significantly lower than the second and third year student groups (85.7 and 88.2 points, respectively). First year students had the lowest ICC and highest MDC (0.23 and 40.4 points). Third year students had the highest ICC and lowest MDC (0.73 and 23.0 points).

Limitations: We were unable to analyze scores from the 13-item MFT for all three student groups. The rater did not receive training in the MFT scoring rubric. Participation in the study was not a requirement.

Conclusions: The 13 and 11-item MFT has good test-retest reliability for UNE's third year DPT student group. The 11-item MFT has poor to moderate test-retest reliability for first and second year DPT students.

Introduction:

According to Sackett et al[1], evidence-based medicine is the thorough and judicious use of the best current evidence in decision making about the care of individual patients. Although beginning in the field of medicine, using evidence in clinical decision making is evolving to include physical therapy and other health care professions. Since this notion is incorporated into practice by all professions, evidence-based medicine is now referred to as evidence-based practice (EBP). Professional organizations, such as the American Physical Therapy Association (APTA), have identified the need for increased training in evidence based practice for their healthcare professions and at every level of education.[2] The APTA has set forth guidelines for doctor of physical therapy (DPT) curricula to educate practitioners who are efficient and critical users of current evidence and who understand how to combine that evidence with their own clinical knowledge and with patients' preferences. The goal is that current practitioners will be lifelong learners who possess the knowledge and skill-set required to remain current with physical therapy's evolving knowledge base.[3]

One method to introduce and teach EBP is by using a five-step framework, referred to as the five A's: ask, acquire, appraise, apply, and assess. The practitioner first asks a focused clinical question which is followed by acquiring the best available current evidence. The evidence should then be critically appraised and applied to the patient population. The final step requires assessment of the outcomes. The Section on Research of APTA recommends DPT education programs use this framework to educate physical therapists how to discover, evaluate, and integrate trustworthy evidence into their professional practice.[3]

Now that DPT programs are teaching evidence-based practice, educators need an assessment tool supported by evidence to evaluate the competence of students.[2] The original Fresno Test (FT) was established to assess EBP knowledge and skill of medical students. Ramos [4] developed the FT for family practice residents at University of California, San Francisco's Fresno family practice residency program. The FT used opened ended questions, some that required calculations, to allow for higher order thinking. Although Ramos [4] found the FT was found to be simple, reliable, and valid for assessment of knowledge and skill, it does not evaluate how the practitioner applies or assesses the evidence.[5] Additionally, the FT is specific to medical students.

Since EBP is utilized by other healthcare professions, the Adapted Fresno Test (AFT) was designed to assess knowledge of EBP in occupational therapists (OT) after a weekend continuing education course. McKluskey [6] adapted the FT to include scenarios specific to OT and found that the AFT was a valid assessment tool. After the FT was successfully adapted for OT, Tilson [5] developed the Modified Fresno Test (MFT) for use with physical therapists. The FT was expanded to encompass PT specific scenarios. She validated it by testing EBP-novice PT students, EBP-trained PT students, and EBP-expert PT faculty. Miller, Cummings and Tomlinson[7] found that the MFT can be scored reliably by trained raters using the rubric made by Tilson.[5] She reported an intra-rater reliability (ICC 2,1) of 0.85 (95% CI 0.60-0.97), an inter-rater reliability of 0.83 (95% CI 0.74-0.96), and a test-retest reliability of 0.46 (95% CI 0.16 - 0.69). She found a minimal detectable change (MDC) of 25.6 points, a value which is used to assess true change in individual performance.[7]

The MDC and test-retest reliability have been determined for 1st year DPT students. However, these values may differ with different levels of EBP knowledge. Therefore, the primary purpose of this study is to determine the test-retest reliability and minimal detectable change of the Modified Fresno Test in 1st, 2nd, and 3rd year DPT students. A secondary purpose is to compare mean MFT scores among the three student groups.

Methods:

Design Overview:

The study used exploratory research to investigate the psychometric properties of test- retest reliability and minimal detectable change.

Setting and Participants:

Using a simple random sample, we recruited 21 UNE DPT students each from the first, second and third year classes. Each of the three classes was represented by the recruited student group. According to Bonett [8], a sample size of 21 subjects will be adequate to detect an intraclass correlation coefficient of 0.90 at alpha of 0.05 with a confidence interval 0.20.

To recruit subjects, we gave a group presentation to each of the three classes describing details of our study. Using a random number generator, we randomly selected 21 students from each class. The selected students were sent an email informing them that they have been selected. We asked the students to reply within 24 hours with their agreement to participate. If they did not wish to participate or did not respond in 24 hours, we selected another student at random until we

had 21 participants. Inclusion criteria included enrollment in the UNE DPT program. There were no exclusion criteria except the investigators of this study (LG and EP).

The study took place at the UNE campus in Portland, Maine. UNE is a private school that enrolls approximately 60 students per DPT class to their eight semester long program (Appendix 2. Curriculum). Students take PTH 514 Scientific Inquiry 1 in the spring semester of their first year and PTH 602 Scientific Inquiry 2 during the fall semester of their second year. These two courses include an introduction to and application of EBP, respectively (Appendix 3. Course Descriptions). At the time of test administrations, first year students had not taken either Scientific Inquiry class; second year students had completed PTH 514 Scientific Inquiry 1; and third year students had completed both Scientific Inquiry courses. Although these two classes directly pertain to EBP, EBP is threaded throughout all three years via assignments in various classes.

MFT Protocol:

The MFT contains 13 questions: eight are short answer response, two require mathematical calculations, and three are fill in the blank response (Appendix 4. Modified Fresno Test). The test takers are introduced to a definition of EBP and are then given two clinical scenarios of three to four sentences each and choose one scenario to answer the 13 items on the test. In general, each response is graded as excellent, strong, limited, minimal, or not evident. Points are assigned to each question based on their pertinence.[5] The total score ranges from 0 - 224, with higher scores suggesting higher EBP knowledge.

The MFT was administered in writing twice to each class of students. We recruited and administered the two tests to second year students near the beginning of the summer semester in June 2014. We recruited and administered the tests to the first and third year students at the beginning of the fall semester 2014 (Appendix 2. Curriculum). Time between the two test administrations of the MFT was 14 days as Miller, Cummings and Tomlinson[7] used in their study. Participants were allowed up to 60 minutes to complete the test in a designated room with supervision. No external resources were permitted except for a calculator. Note paper was provided. Each participant was assigned a separate identification number for each test in order to mask the participants identity and version of the MFT (test one and test two) to prevent grading bias. After both tests were completed they were scored on paper by Michael Fillyaw PT, MS, an EBP expert, professor of PTH 514 Scientific Inquiry 1 and our research advisor for this study, using Tilson's scoring rubric[5] (Appendix 4. Modified Fresno Test).

Statistical Analysis:

To achieve the primary purpose of the study, test-retest reliability was calculated separately for MFT total score for each student group using intraclass correlation coefficient [ICC (2,1)].[9] Standard error of measure (SEM) was calculated separately for MFT total score for each of the three student groups using: $SEM = SD_{initial}\sqrt{1 - ICC}$ where $SD_{initial}$ is the standard deviation of the scores for test one.[10] The SEM was then used to calculate the MDC for each student group using: $MDC_{95} = \text{Critical Z-Score}_{95} * SEM * \sqrt{2}$ where the critical z-score for 95% level of confidence is 1.96.[10]

To achieve the secondary purpose, we used an analysis of variance to test for the significance of the difference in mean total scores among the three groups. We also tested for the significance of the difference in mean total scores for test one and test two within the three student groups by a paired t-test. A level of significance of p-value <0.05 was used for both analyses. SYSTAT 13 was utilized for all of our statistical analyses.

Results:

A total of 53 students, 20 from the third year class, 18 from the second year class and 15 from class the first year class, who participated in both test administrations, were analyzed for the primary and secondary purposes of our study (Figure 1).

Third year students received the complete 13-item MFT as intended. Due to a photocopying error, questions 10 and 11 were not included in the test packets received by first and second year student groups, therefore, first and second year students did not complete the 13-item MFT as designed. We will refer to the MFT without questions 10 and 11 as the "11-item MFT".

To achieve our secondary purpose, the mean and standard deviation for total score for the 13-item MFT were calculated only for the third year students. The mean and standard deviation for total score of the 11-item MFT were calculated for each of the three classes (Table 1). The ANOVA showed significant differences in mean total scores for the 11-item MFT among three student groups (Table 2). The third year student group had the highest mean score and the first year student group had the lowest. A Bonferroni post hoc analysis was done to determine which pairs were different among the three classes. This analysis revealed that the second and third year student groups were different from the first year student group (Table 2 and Figure 2).

To achieve our primary purpose, the ICC and MDC were calculated for the total score of the 13-item MFT for the third year students and for the 11-item MFT for each student group (Figures 3 and 4). The third year student group had the highest ICC values of 0.77 and 0.73 for their 13-item and 11-item MFT, respectively (Figure 3). The third year student group also had the lowest MDC values of 21.61 and 23.04 points for the 13-item and 11-item MFT, respectively (Figure 4). The first year students had the lowest ICC value of 0.23 and the highest MDC value of 40.4 points (Figure 3 and 4).

The second year students were recruited in May of 2014 and completed both test administrations in June of 2014. The first and third year students were recruited in August of 2014. Test one was completed at the end of September of 2014 and test two was completed in the beginning of October of 2014. No adverse events took place during the study.

Discussion:

We suspected that mean MFT scores would increase with increased exposure to EBP. The results of the Bonferonni post hoc analysis supported our hypothesis. The analysis found a significantly lower mean total score on the 11-item MFT in the first year students compared to both the second and third year students. The difference in mean total scores of the 11-item MFT between the second and third year classes was not statistically significant, however the third year class scored higher than the second year class, as expected. This difference may not be significant because third year students may have given less effort than second year students. The third year students may have felt obligated to participate in order to assist fellow classmates (investigators EP and LG). Lack of motivation was noted in this student group, especially during

159 the second test administration. Students in the first and second year classes were supervised by a
160 faculty or staff member, whereas students in the third year class were supervised by peers.
161
162 We suspected there would be a difference in ICC and MDC values among each group of
163 students. We found that higher ICC values were associated with a higher total MFT score. None
164 of our ICC values indicate high test-retest reliability (>0.90), however the ICC value for the third
165 year class (0.73) allows us to consider the 11-item MFT as having good reliability for this
166 student group.[9] The first and second year classes revealed ICC values (0.63 and 0.23) indicative
167 of poor to moderate test-retest reliability.[9]
168
169 The ICC is determined both by the extent of agreement in participant scores as well as variability
170 of their scores. ICC will be highest when scores agree from test one to test two and there is a
171 heterogeneous group of participants.[9] Unlike the second and third year classes, at the time of test
172 administrations the first year students had no exposure to EBP through UNE's DPT curriculum.
173 Any experience the first year students had would be unique to each participant, resulting in more
174 variability in their scores than second and third year classes. Because of this, we initially
175 suspected that the first year students would have a high test-retest reliability. However, our first
176 year class was the most homogenous group with respect to MFT total score. The students' mean
177 total scores were very similar to one another, and revealed minimal knowledge of EBP, having
178 the lowest mean score out of all three student groups. The third year student group was the only
179 group to have a significant difference in mean total scores on the 11-item MFT from test one to
180 test two (Figure 5), but their MFT total scores were the most variable. The large amount of
181 variability among students was sufficient to overcome the significant difference in mean total
182 scores and resulted in the highest ICC value. The variability in scores of the third year class is

understandable because although the students had been exposed to EBP through the curriculum, each student had established a unique and variable amount of EBP knowledge.

The option to choose one of two scenarios for the MFT may have influenced the test-retest reliability through decreasing the agreement of scores. There was no prohibition against students choosing one scenario on test one and choosing the other scenario on test two. Although both scenarios were considered valid for the physical therapy population[5], if a participant did not choose the same scenario on both test administrations, he/she could have essentially taken two different tests.

The open-ended responses required on the MFT may have contributed to lower test-retest reliability compared to another more objective form of testing. Students could have easily and unknowingly altered their responses from test one to test two. The open-ended format could have also impacted scores due to the rater. Scoring the open-ended responses relied heavily on the rater's interpretation of answers, which may not have been consistent within or among subjects.

The MDC values also differed between classes. Students in the third year class had the highest test-retest reliability and the lowest MDC value (23.04 points). The MDC value for the second year class (27.15 points) was similar to that of the third years and the ICC values were also similar. The first year class had the lowest ICC value (0.23) and the highest MDC value (40.38 points).

To our knowledge, this is the first study to examine the test-retest reliability of the validated 13-item MFT and our 11-item MFT. Miller, Cummings and Tomlinson[7] looked at the test-retest

reliability and total mean score of the original 14-item MFT for first year DPT students and found an ICC of 0.46 with a MDC of 18.2 points and mean total score of 107 points from a possible total score of 232 points. In her validity study, Tilson[5] found that question nine in the 14-item MFT exhibited unsatisfactory psychometric properties and recommended not including it in the test. The third-year students was the only group in our study to take the validated 13-item MFT (total possible points 224). The mean score of the 13-item MFT for the third year student group was 88.2 with a MDC of 21.61 and ICC of 0.77.

This is the first study of the MFT to utilize students who have not received formal EBP training through a DPT program. Miller, Cummings and Tomlinson's[7] novice group received a short EBP course consisting of seven to ten hours of EBP content. This was similar to Tilson's study whose novice group received seven hours of EBP training which included an introduction to EBP, writing searchable questions, searching, and appraising literature. The participants from our first year class, those with the least amount of EBP exposure, did not receive an introductory course to EBP as a part of this study. Our findings agree with Tilson's[5], that novice evidence based practitioners had the lowest MFT scores and most experienced evidence based practitioners had the highest MFT scores.

Tilson[5] weighted questions one through seven higher (worth 24 points each) than other questions (ranging from 4 to 16 points) since they were considered to have greater importance. Our third year students scored higher on these questions than our first and second students. Additionally the third year students scored low on questions 10 and 11 which asked about relative risk, likelihood ratio, number needed to treat, positive predictive value, risk reduction, sensitivity, or P-value. This low score may be due to little emphasis UNE's DPT program places on these

mathematical calculations. The inclusion of questions 10 and 11 resulted in a mean score of 5.7 points higher for the 13-item MFT compared to the 11-item MFT when there was a possibility to score an additional 28 points. Knowing the scores for each individual question allows faculty at UNE to identify areas of strength and weakness in the EBP curriculum and to make changes as desired.

Limitations:

Due to a photocopying error the statistical analysis was limited to comparing the 11-item MFT for the classes of 2016 and 2017. This did not allow us to analyze the ICC, MDC, or total mean score for the 13-item MFT for the classes of 2016 and 2017.

The rater of the tests did not receive formal training in the scoring rubric of the MFT as was provided for the raters in other studies[5,7](Appendix 4. Modified Fresno Test). The MFT test-retest reliability also depends on consistency of raters scoring the test for accuracy and there may have been variance within our rater's interpretation of the student's answers.

Participation in the study was not required and, therefore, participants may not have taken it seriously resulting in lower and varied scores from test one and test two.

Conclusions:

The 13-item and 11-item MFT can be considered as having good test-retest reliability for a third year DPT student group with similar EBP exposure as taught in UNE's DPT program. The 11-item MFT exhibited poor to moderate test-retest reliability for first and second-year DPT students. The high MDC value in the first year student group, which was associated with their low test-retest reliability, highlights the need for a first year DPT student to score 40.4 points

higher or lower to demonstrate meaningful change. Future research is needed to assess the test-retest reliability of the validated 13-item MFT for students in the first and second year class of a DPT program.

References

1. Sackett DL, et.al: Evidence-based medicine: what is it and what isn't it? *BMJ* 1996; 312:71
2. Shaneyfelt T, Baum KD, Bell D, et al. Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA* 2006; 296 (9): 1116-1127.
3. APTA Evidence-Based Practice Special Interest Group Section on Research. *Evidence-Based Practice Curriculum Guidelines for Doctor of Physical Therapy Education.* Alexandria, VA: American Physical Therapy Association; 2014.
4. Ramos KD, Schafer S, Tracz SM: Validation of the Fresno test of competence in evidence based medicine. *BMJ (Clinical research ed)* 2003; 326(7384):319.
5. Tilson JK. Validation of the modified Fresno test: assessing physical therapists' evidence based practice knowledge and skills. *BMC Med Educ* 2010; 10:38.
6. McKluskey A, Bishop B: The Adapted Fresno Test of competence in evidence-based practice. *J Contin Educ Health Prof* 2009, 29 (2): 119-126.
7. Miller AH, Cummings N, Tomlinson J. Measurement Error and Detectable Change for the Modified Fresno Test in First-year Entry-level Physical Therapy Students. *J Allied Health* 2013; 42 (3): 169-174.
8. Bonnet DG. Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in Medicine*. 2002; 21:1331-1335.
9. Portney LC, Watkins MP. *Foundations of Clinical Research: Application to Practice*. 2nd ed. Upper Saddle River, NJ: Prentice Hall Health; 2000.
10. DiFabio R. *Essentials of Rehabilitation Research: A Statistical Guide to Clinical Practice*. Philadelphia, PA; FA Davis Company; 2013.

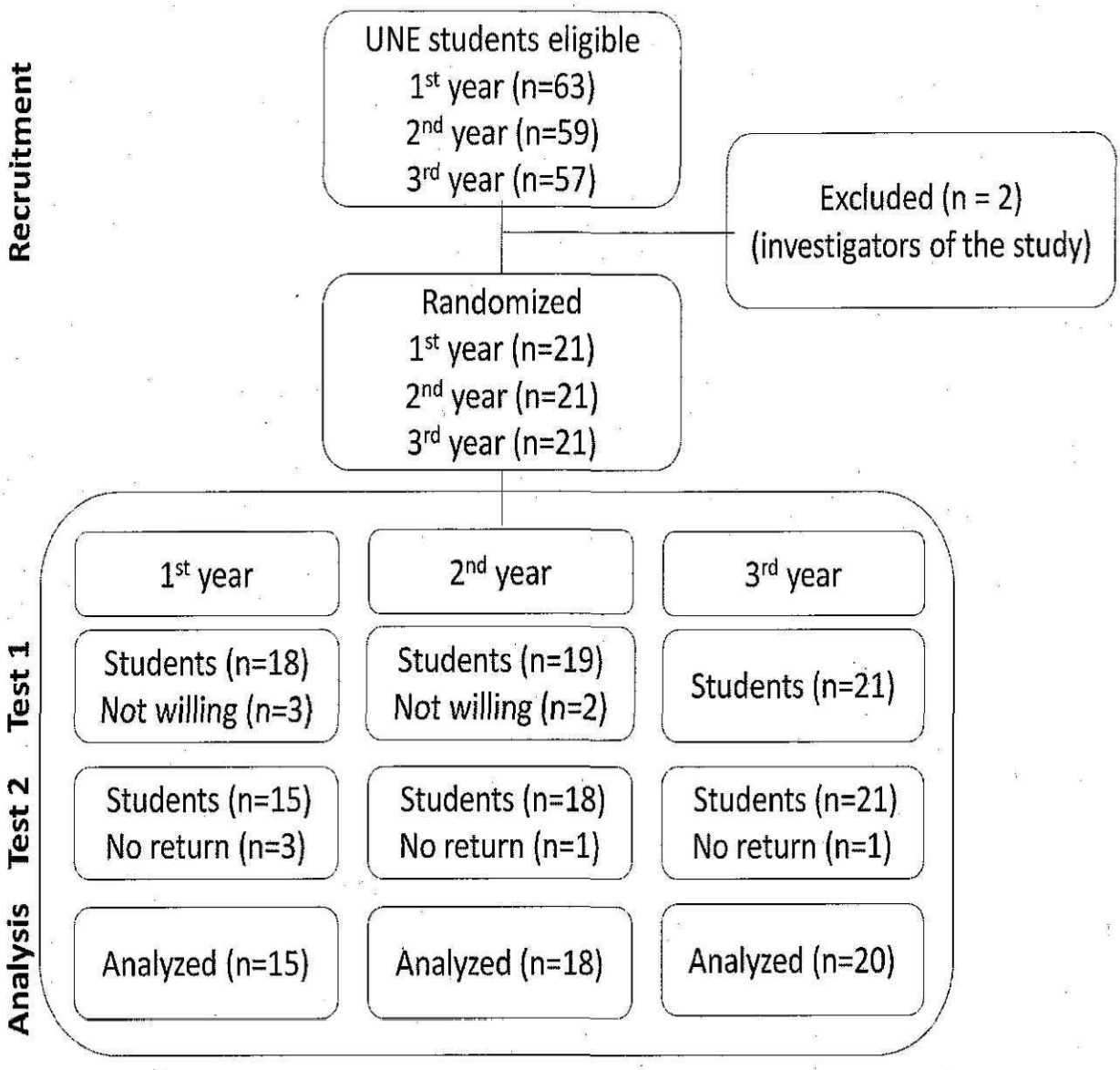

Figure 1 MFT Participation Flow Chart 21 students from each DPT class were recruited. After both test administrations, we were able to analyze the MFT scores for 53 students.

Table 1 Differences between Test 1 and Test 2 for each class

	First Year					Second Year					Third Year				
	T1		T2		P-value	T1		T2		P-value	T1		T2		P-value
MFT Question (Possible Points)	Mean	SD	Mean	SD		Mean	SD	Mean	SD		Mean	SD	Mean	SD	
1. Form a clinical question (24)	11.1	5.1	12.8	4.8	>0.05	18.8	3.8	18.1	4.4	>0.05	20.2	3.7	18.2	4.7	>0.05
2. Sources of information (24)	15.5	6.6	13.1	4.9	<0.05	8	5.4	6.6	5.1	>0.05	9.6	4.5	7.4	4.9	>0.05
3. Study design (24)	8.6	6.9	6.9	5.8	>0.05	13.2	4.7	13.5	4.7	>0.05	16.7	5.4	12	5	<0.05
4. Search strategy (24)	8.2	2.8	8.7	3.6	>0.05	18.3	5.2	16.6	6.1	>0.05	16.3	4.7	15.6	5.1	>0.05
5. Relevance of study (24)	7.5	5.9	10.8	5.2	>0.05	12.4	5.8	9.9	4.5	>0.05	11.8	4.9	12	7.2	>0.05
6. Validity of study (24)	7	6.2	4.4	4.4	<0.05	8.2	6.8	5.9	5.3	>0.05	4	4.2	3.9	5.2	>0.05
7. Magnitude, significance of study (24)	3.9	2.8	3.6	3.2	>0.05	5.9	5.6	7.5	4.4	>0.05	7.6	4.8	7	4.9	>0.05
8. Questions for patient/family (8)	3.4	2.3	2.2	1	<0.05	2.6	1.5	3.1	1.6	>0.05	3	2	2.4	1	>0.05
10. Sensitivity, PPV, LR (12)	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	1.4	2.3	1.3	1.8	>0.05
11. RR, NNT, p-value (16)	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	3.5	2.1	5.4	2.6	<0.05
12. Confidence Interval (4)	0.0	0.0	0.0	0.0	>0.05	0.9	1.7	0.9	1.7	>0.05	1	1.8	2.4	2	<0.05
13. Best study design - diagnosis (4)	0.3	1	0.6	1.4	>0.05	0.2	0.9	0.2	0.9	>0.05	0.4	1.2	1.6	2	<0.05
14. Best study design - prognosis (4)	1	1.7	0.6	1.4	>0.05	0.4	1.3	0.2	0.9	>0.05	2	2.1	1.4	2	>0.05
MFT total 13-item (224)	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	97.3	16.3	90.55	21.4	>0.05
MFT total 11-item (196)	71.8	16.6	65.3	15.1	>0.05	88.9	16.1	82.5	12.4	>0.05	92.5	16.0	83.9	20.2	<0.05

MFT= Modified Fresno Test
T1= MFT test administration 1
T2= MFT test administration 2
SD= standard deviation
NT= not tested
PPV=positive predictive value, LR= likelihood ratio
RR= risk reduction, NNT= number needed to treat
MFT total 13-item= sum of questions 1-8, 10-14
MFT total 11-item= sum of questions 1-8, 12-14

Table 2. Comparison of each question and total score among classes

	1st year		2nd year		3rd year		P-value	Bonferroni Test
MFT Question (Possible Points)	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**		
1. Form a clinical question (24)	12.8	4.4	18.4	4.1	19.2	4.3	0.000	2015≠ 2017; 2016≠2017
2. Sources of information (24)	15.3	5.0	7.3	5.2	8.5	4.8	0.000	2015≠ 2017; 2016≠2017
3. Study design (24)	8.1	6.5	13.3	4.7	14.3	5.6	0.000	2015≠ 2017; 2016≠2017
4. Search strategy (24)	8.9	3.0	17.4	5.7	15.9	4.8	0.000	2015≠ 2017; 2016≠2017
5. Relevance of study (24)	9.6	5.9	11.2	5.3	11.9	6.1	0.274	No differences
6. Validity of study (24)	6.0	5.8	7.1	6.1	4.0	4.7	0.049	2015≠ 2016
7. Magnitude, significance of study (24)	3.8	3.0	6.7	5.0	7.3	4.8	0.004	2015≠ 2017; 2016≠2017
8. Questions for patient/family (8)	2.9	1.9	2.8	1.5	2.7	1.6	0.845	No differences
10. Sensitivity, PPV, LR (12)	NT	NT	NT	NT	1.3	2.0	NT	NT
11. RR, NNT, p-value (16)	NT	NT	NT	NT	4.5	2.5	NT	NT
12. Confidence Interval (4)	0.0	0.0	0.9	1.7	1.7	2.0	0.000	2015≠ 2017
13. Best study design - diagnosis (4)	0.4	1.2	0.2	0.9	1.0	1.8	0.039	2015≠ 2016
14. Best study design - prognosis (4)	0.8	1.6	0.3	1.1	1.7	2.0	0.002	2015≠ 2016
MFT total 13-item (224)	NT	NT	NT	NT	93.9	19.1	NT	NT
MFT total 11-item (196)	68.5	15.9	85.7	14.5	88.2	18.5	0.000	2015≠ 2017; 2016≠2017

MFT= Modified Fresno Test, MFT total 13-item= sum of questions 1-8, 10-14, MFT total 11-item= sum of questions 1-8, 12-14
SD= standard deviation
PPV=positive predictive value, LR= likelihood ratio
RR= risk reduction, NNT=number needed to treat
NT= not tested

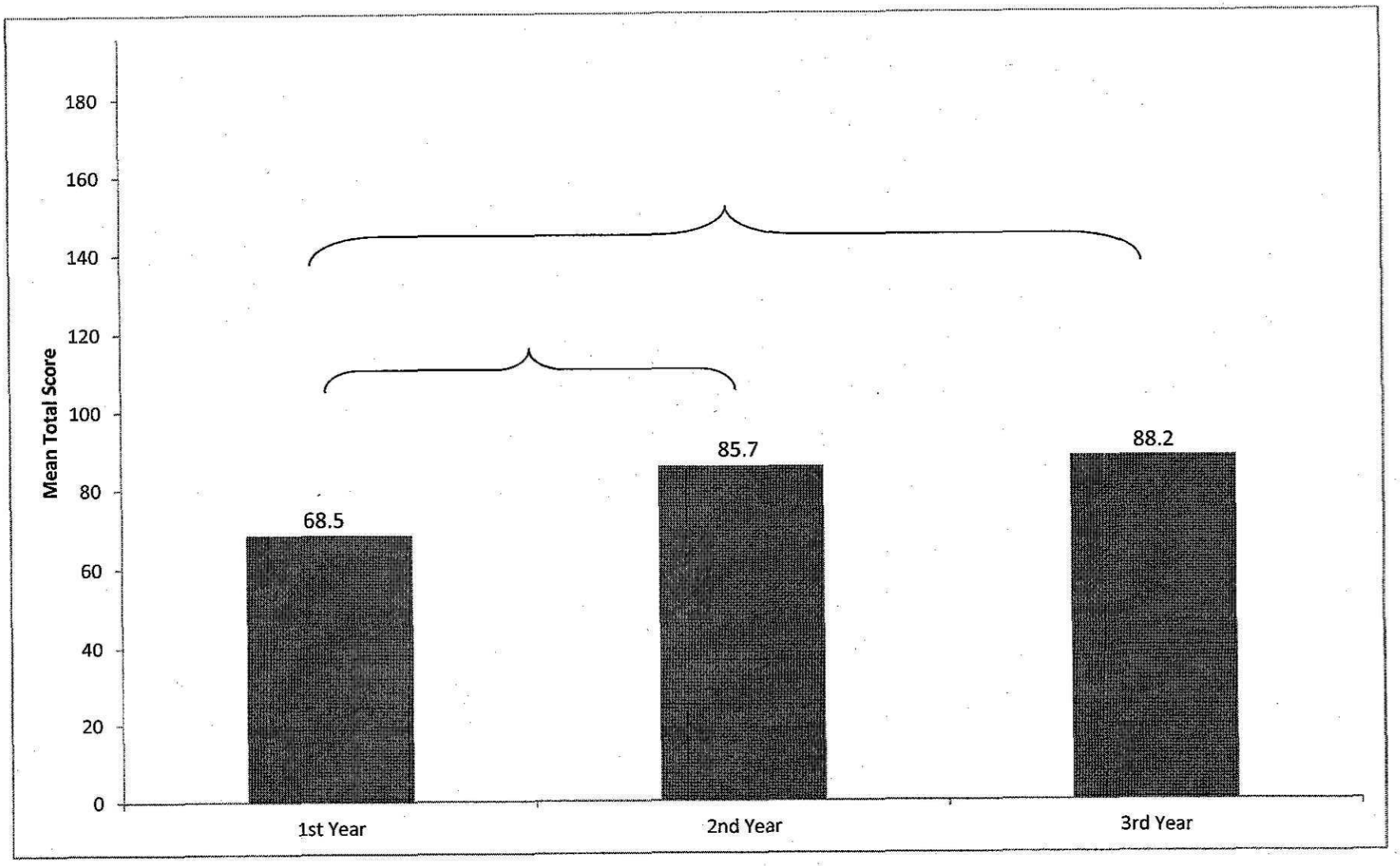

Figure 2 Mean Total Scores Brackets indicate a significant difference in mean total 11-item MFT scores between 1^{st} & 3^{rd} and 1^{st}& 2^{nd} year students.

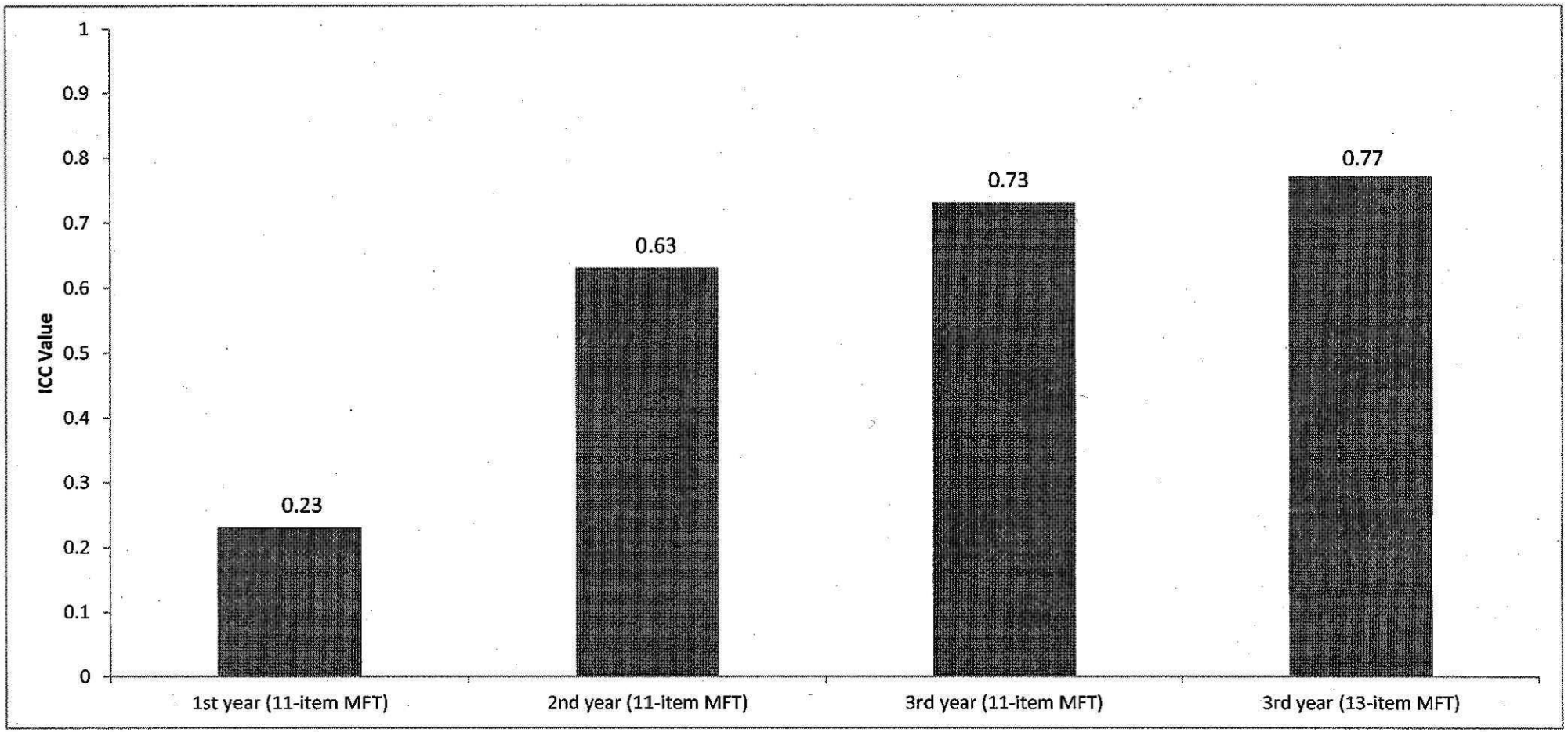

Figure 3 Test-Retest Reliability Increasing ICC values are associated with increasing exposure to EBP. The 13-item MFT given only to 3rd year students had the highest ICC value.

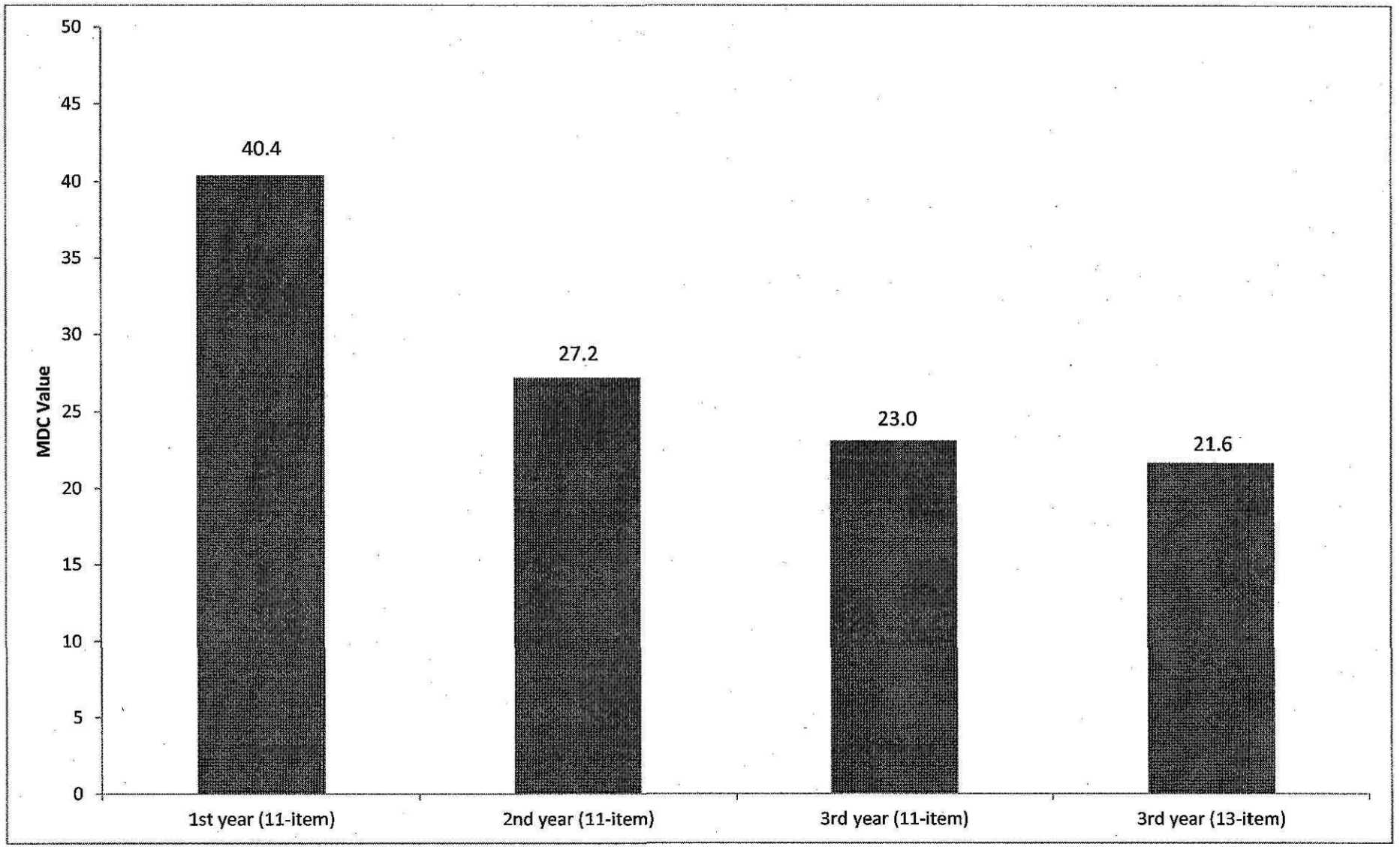

Figure 4 Minimal Detectable Change The 1st year class must score 40.4 points higher on subsequent 11-item MFT to demonstrate a true change in scores which is much higher than the 2nd and 3rd year classes. A higher MDC is associated with lower exposure to EBP and a lower MFT score.

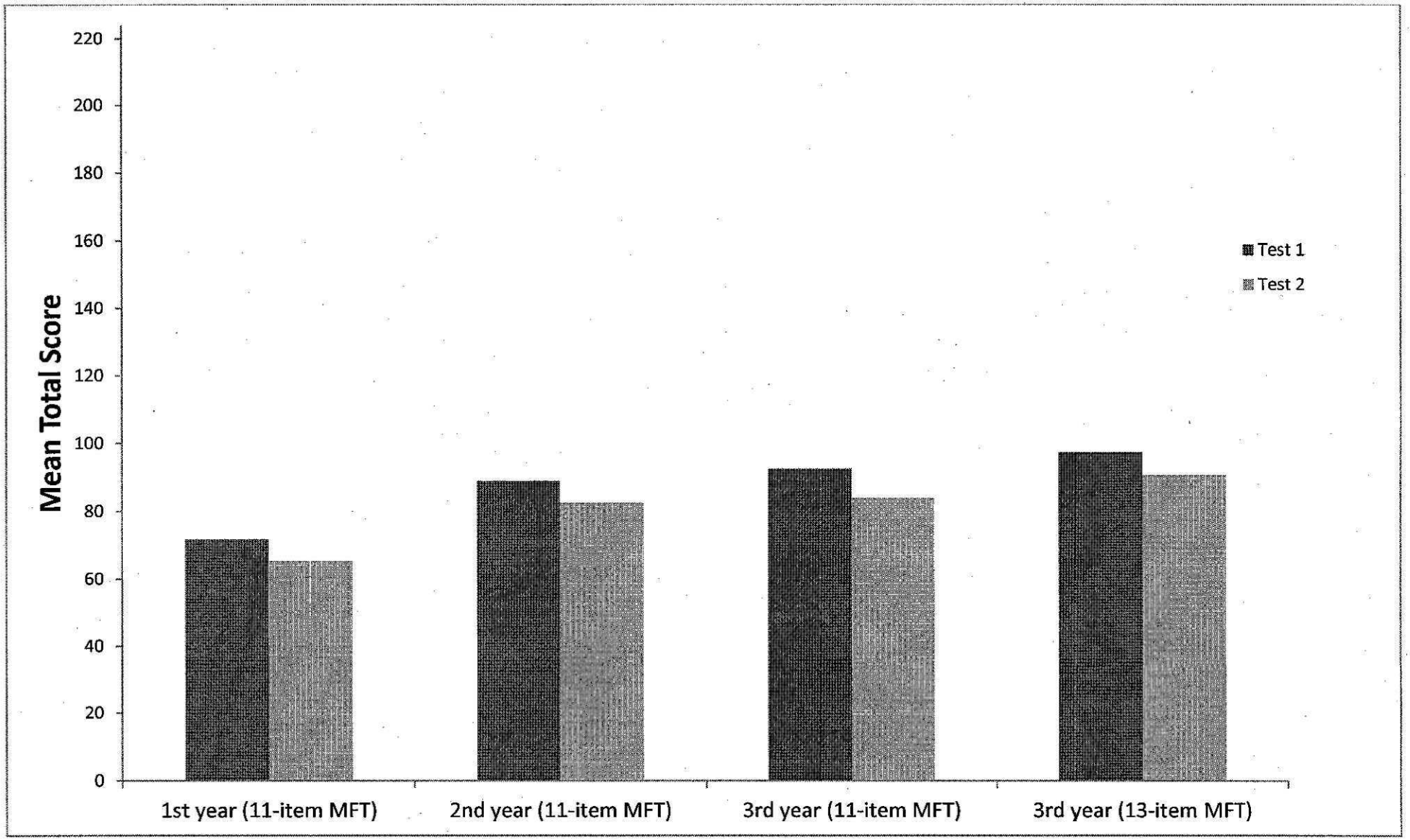

Figure 5 Mean Total Score 11-item MFT mean total scores Test 1, Test 2: 1st years 71.8, 65.3; 2nd years 88.9, 82.5; 3rd years 97.3, 90.6. 13-item MFT mean total scores Test 1, Test 2: 92.5, 88.3. The only statistically significant difference in mean total scores between test 1 and test 2 was on the 11-item MFT for third year students. Total possible points for the 11-item MFT is 196 points. Total possible points for the 13-item MFT is 224 points.

Appendix 1. Budget and Budget Justification

RESEARCH PROJECT BUDGET			Period from: May 2014 Through: December 2014	
A. All personnel for whom money is requested	Annual Salary	Benefits	% time on project	Requested funds
Erin Pike	$65,000	$16,250	12.5%	$10,156.25
Lisa Gerhardt	$65,000	$16,250	12.5%	$10,156.25
			Subtotal A =	$20,312.50
B. Equipment				Requested funds
			Subtotal B =	$0.00
C. Consumable Supplies				Requested funds
Paper				$34.02
Pencils				$13.99
			Subtotal C =	$48.01
D. Dissemination				Requested funds
Student CSM registration				$280.00
Airfare				$610.50
Hotel				$376.00
Meals and Incidental expenses				$355.00
			Subtotal D =	$1,621.50

Appendix 2. UNE DPT Curriculum

CURRICULUM BY SEMESTER

YEAR 1		YEAR 2		YEAR 3	
		AY2 – Summer	**11**	**AY3 – Summer**	**8-10**
		BIO 504 – Neuroscience	4	PTH 607 – Clinical Practicum 2	8
		PTH 516 – Pathology & Med Management: Cardiovascular	1	PTH 608 – Case Report 1 (Distance Education Course)	2*
		PTH 522 – PT Management of Patients – Disorders: Cardiovascular and Pulmonary Systems	4		
		PTH 525 – Practice Management	1		
		PTH 524 – Clinical Education Seminar	1		
AY1 Fall	**17**	**AY2 – Fall**	**10**	**AY3 – Fall**	**13**
BIO 502 – Gross Anatomy	6	PTH 601 – Clinical Practicum 1	8	PTH 700 – Administration	2
PTH 501 – Foundations of PT Practice	5	PTH 602 – Scientific Inquiry 2 (Distance Education Course)	2	PTH 701 – Pathology & Med Management: Disorders of the Integumentary System	1
PTH 502 – Kinesiology	5			PTH 703 – PT Management of Patients: Disorders Of the Integumentary System	4
PTH 507 – Introduction to Clinical Medicine	1			PTH 704 – Disease Prevention & Health Promotion	3
				PTH 705 – Research Project	2*
				PTH 708 – Case Report 2	2*
				PTH 710 – Complex Case Management	1
AY1 – Spring	**18**	**AY2 - Spring**	**14-16**	**AY3 – Spring**	**10**
PTH 503 – Normal Development	2	PTH 603 – Pathology & Med Management: Disorders of the Neuromuscular System	3	PTH 706 – Public Policy and Physical Therapy	2
PTH 506 – Psychosocial Aspects of Disability and Illness	1	PTH 604 – PT Management of Children: Disorders of the Neuromuscular System	5	PTH 707 – Clinical Practicum 3	8
PTH 508 – Pathology & Med Management: Musculoskeletal System	2	PTH 605 – PT Management of Adults: Disorders of the Neuromuscular System	6		
PTH 510- PT Management of Patients – Disorder: Musculoskeletal System	11	PTH 606 – Research Proposal	2*		
PTH 514 – Scientific Inquiry 1	2				

Appendix 3. Course Descriptions

PTH 514 Scientific Inquiry 1: An introduction to the methods used to conduct clinical research in physical therapy. Topics include: accessing the professional literature using electronic databases, evidence-based practice, and evaluating the literature of physical therapy through examination of the research process including sampling, experimental design and control, ethics of clinical research, properties of measurements, and statistical inference.

PTH 602 Scientific Inquiry 2: The physical therapy student applies the principles of evidence-based practice to clinical problems. This includes: 1) asking patient-centered questions, 2) identifying, searching, and critically appraising published sources of evidence, and 3) integrating the evidence along with clinical expertise, and the patient's circumstances and preferences into clinical decisions. This course is offered in distance-learning format concurrent with PTH 601: Clinical Practicum I. Students are required to have a computer with Microsoft Word and Internet access.

Appendix 4. Modified Fresno Test and scoring rubric

Instructions:
Evidence Based Practice (EBP) involves knowledge and skills related to identifying and evaluating evidence to inform practice. This tool, the modified Fresno Test is designed to assess your EBP skills.

There are 8 short answer questions, 2 questions that require a series of mathematical calculations, and 3 fill-in-the-blank questions. A calculator and note paper have been provided for you. Additional resources (internet sites, books, etc) are not permitted.

Please complete the entire test in one sitting and allow yourself up to 60 minutes to complete the test.

Answer questions 1-4 and 8 based on the following clinical scenarios:

Scenario 1: You have just evaluated Mary, a secretary who recently experienced a work related low-back injury moving 10, 25 lbs. file boxes 3 days ago. Her radiographs are negative and her only symptom is resolving 2/10 pain across the low back with forward bending and prolonged sitting. She has been off of work for 2 days and is eager to return but is also anxious about re-injury. You are considering a stabilization exercise program but wonder if manual therapy should be included in the patient's physical therapy program.

Scenario #2: Marvin is a 10 year-old boy with hemiparesis secondary to stroke associated with an Arterial-Venous Malformation. He presents to outpatient therapy and his parents express particular concern about Marvin's arm and leg weakness. You are considering implementing an intensive task-specific strengthening program but a colleague warns that such a program is likely to increase the patient's moderate flexor tone and spasticity and suggests low intensity stretching and a passive positioning program.

QUESTION #1: Choose one of the above clinical scenarios. Write a focused clinical question for that scenario that will help you organize a search of the clinical literature.					
	Excellent	Strong	Limited	Minimal	Not evident
a: Population	6: Multiple relevant descriptors; e.g., "work-related injury," "female" or "acute," or " or "low-back pain"; e.g., "boy with hemiparesis" specific age group, gender, diagnosis, motor presentation	4: One appropriate descriptor as above examples; e.g. "women" or "worker" or "low-back pain"; e.g. "hemiparesis" "boy" "10 year old" "post-stroke"	2: A single general descriptor unlikely to contribute to search; e.g. "patient"		0: None of the above present
b: Intervention	6: Includes specific intervention of interest; (intervention could be a diagnostic technique); • manual therapy • specific individual components of manual therapy • combination of exercise and manual therapy; • task-specific strengthening		2: Mentions intervention but unlikely to contribute to search; e.g. "methods" "options" "treatments"		0: None of the above present
c: Comparison	6: Identifies specific alternative of interest; e.g. "no manual therapy"; "low intensity stretching"		2: Mentions comparison but unlikely to contribute to search; e.g. "alternate methods"		0: None of the above present
d: Outcome	6: Outcome that is objective and meaningful to patient or patient case (if question is diagnostic, should relate to diagnosis trying to detect); e.g.return to work, pain reduction, injury prevention; e.g. selective motor control or functional use of paretic extremities, walking velocity	4: Non-specific outcome: • recovery • spasticity • tone • strength	2: Reference to outcome, but so general as to be unlikely to contribute to search • effects • change the outcome • effective • improvement • success • change the outcome		0: None of the above present

QUESTION #2: Where might you find answers to this and other similar clinical questions? Name as many possible sources of information as you can – not just the ones you think are 'good' sources. Describe the advantages and disadvantages of each type of information source you have listed.					
	Excellent	Strong	Limited	Min	Not evident
a: Variety of Sources	6: At least four types of sources listed. Types include: • electronic databases of original literature [Medline(PubMed/Ovid), CINAHL] • discipline specific databases (Hooked on Evidence, PEDro) • journals (JAMA, NEJM, incl. access through library) • text book (Merck, Harrisons, monographs) • Systematic Reviews (Cochrane) • EBM publications or databases of pre-appraised information (Best Evidence, InfoRetriever, DynaMed, EBM, ACPJC, EBP, Clinical Evidence) • Medical website (MDConsult, PraxisMD, SumSearch) • General internet search (google, yahoo) • Clinical Guidelines (Guideline Clearinghouse, • Professional Organization (AAFP, La Leche League, NIH website) • People (colleague, consultant, attending, librarian)	4: Three types of sources listed.	2: Two types of sources listed.		0: No variety. Only one source listed, or all sources of same type.
b: Convenience	6: Discussion includes at least 2 specific issues related to convenience, or mentions the same issue while discussing two different sources. Issues may include: • Cost (e.g. "free," "subscription only") • Speed (e.g. "fast," "takes time") • Ease of search (e.g. "must know how to narrow search," "easy to navigate") • Ease of use (e.g. "concise" and "NNTs already calculated") • Availability (e.g. "readily available online")	4: Includes 1 specific issue/explanation related to convenience	2: Mentions convenience involved in using one or more source, but without explanation e.g. "convenient" or "easy" or "difficult"		0: No mention of convenience

#2 Continued

	Excellent	Strong	Limited	Not evident
c: Clinical Relevance	6: Discussion includes at least 2 specific issues related to relevance, or mentions the same issue while discussing two different sources. Issues may include: • Clinically relevant outcomes • Written for clinical application (e.g. "pertinent" "info on adverse effects" or "has patient information sheets") • Appropriate specialty focus (e.g. "directed at PTs") • Information applicable to patient in question (e.g. "can go over details of this particular patient" or "most of studies are from Europe") • Includes specific interventions in question • Specificity (overview vs. targeted) (e.g. "can get basic information" or "more specialized") • Comprehensiveness of source (likelihood of finding an answer in that source) (e.g. "she can find anything" or "contains usable references" or "not likely to have answer to this question")	4: Includes 1 specific issue/explanation related to relevance	2: Mentions relevance of using one or more source, but without explanation e.g. "relevant"	0: No mention of relevance
d: Validity	6: Discussion includes at least 2 specific issues related to validity, or mentions the same issue while discussing two different sources. Issues may include: • Certainty of validity (e.g. quality is uncertain" or "has not been screened" or "needs to be critically appraised") • Evidence Based approach (e.g. "evidence based" or "Grade 1 Evidence" or "no references provided") • Expert bias (e.g. "usually just someone's opinion") • Systematic approach • Peer review • Ability to verify • Standard of care (e.g. "accepted in medical community") • Enough information provided to critique validity (e.g. "abstract only" or "not available full-text") • Up-to-date/outdated (e.g. "most recent research") • Reliability – in the context of the degree of trust that can be places on the resource	4: Includes 1 specific issue/explanation related to validity	2: Mentions validity of using one or more source, but without explanation e.g. "good" "junk"	0: No mention of validity

QUESTION #3: What type of study (study design) would best answer your clinical question outlined in Q1 and why?					
	Excellent	Strong	Limited	Minimal	Not evident
a: Study Design	12: Names one of the best sources: • Randomized Controlled Trial • Randomized Trial • Systematic Review; • Meta-Analysis • Randomized, Double Blinded Clinical Trial	9: Describes but does not call by name one of the best sources as above • comparing two groups, one gets treatment, other gets placeo • double blind study	6: Describes or names a less desirable study design: • Cohort study • Prospective clinical trial • meta-analysis of such studies • Longitudinal or prospective	3: Describes or names a poor study design to answer a treatment question: • case control, cross sectional study, case report, "retrospective" • Or describes a study with insufficient detail to identify a design: e.g. a comparative study	0: None of above present
b: Justification	12: Includes well-reasoned justification that reflects understanding of the importance of randomization and/or blinding. Explicitly connects randomization to reduction of confounding and/or blinding to observer or measurement bias. e.g. "An RCT will attempt to avoid any bias which would influence the outcome of the study through randomization" OR "best suited for therapy questions because it reduces bias and controls for confounding factors."	9: Justification is present, and touches on issues related to randomization and/or blinding, but less clearly articulated e.g. "groups should be similar" or "try to eliminate confounding factors" or "avoid selection bias" or "to be objective" or "to eliminate bias"	6: Justification is present, and raises legitimate issues unrelated to randomization or blinding, such as cost effectiveness, ethical concerns, recall bias. May mention randomization or blinding but without explanation. (e.g. "best in a random and blind setting") e.g. "chart reviews provide lots of data without much cost"	3: Attempted justification, but arguments are non-specific and do not demonstrate understanding of the relationship between the design and various threats to validity May mention randomization or blinding but without explanation. (e.g. "best in a random and blind setting") e.g. "to ensure quality" or "to reduce potential conflicts" or "to compare"	0: None of above present

QUESTION #4: If you were to search Medline, CINAHL or any other database for original research to answer your clinical question related to the scenario you selected for Question 1, describe the search strategy you might use. Be as specific as you can about the search terms and search fields you would use. Explain your rationale for taking this approach. Describe how you might limit your search if necessary and explain your reasoning.					
	Excellent	Strong	Limited	M	Not evident
a: Search Terms	8: 3 or more terms that reflect patient, intervention, comparison, and outcome (PICO) being considered	6: 2 terms from PICO	3: 1 term from PICO		0: Not present
b: Tags/Strategy	8: Description of search strategy reflects understanding that articles in database are indexed by more than one field. Discusses one or more field/index/tag by name (MeSH, Title Word, Publication Title, language, Keyword, author, Journal title, use of boolean operators, etc.) **AND** provides plausible rationale for search strategy using 1 or more of these indices e.g. "keyword is less specific than MESH"	6: Names 1 or more field or index category but does not provide plausible defense of search strategy based on this knowledge e.g."I would do a keyword search…followed by…" "I would use terms … in this way"	3: weak description of strategy, no name given to tags, or overtly misguided strategy e.g. "I would use terms …." [no description of strategy]		0: No evident understanding that articles "tagged" by different fields or indices
c: Delimiters	8: Describes more than one approach to limiting search (e.g., "limit to human" or "adult" or "English"), names a specific publication type, or describes of Clinical Queries in PubMed, or the use of Boolean operators or search combinations or includes terms related to an optimal study design (e.g. randomized) or suggests use of subheadings * NOTE: If the subject includes the name of the index when describing a delimiter (e.g. "check language as English") then we give credit for a tag as well as a method of delimiting.	6: Describes only 1 common method of limiting search e.g. describes ways to narrow search using keywords but no other strategies listed	3: provides weak explanation or description of use of limiters/narrowing search		0: No valid techniques for limiting a search listed

	severity as my patient?" or • "did selection or inappropriate inclusion criteria result in a study population that differs from mine by race, age, etc"		criteria?" or • "selection bias" or • "setting" or • "where study was conducted"	

NOTE: RESPONSES TO QUESTIONS 5, 6, AND 7, CAN BE APPLIED TO ANY PORTION OF THE GRADING RUBRIC FOR THOSE ITEMS

QUESTION #5: When you find a report of original research on this question or any others, what characteristics of the study will you consider to determine if it is relevant? Include examples. Questions 6 and 7 will ask you how to determine if the study is valid, and how important the findings are. For this question, please focus on how to determine if it is really relevant to your practice.				
	Excellent	Strong	Limited	Not evident
a: The Question	12: **Well-reasoned and thoughtful discussion of the relevance of the independent and dependent variables used in the study including examples/specific reasons.** May discuss (well-reasoned and thoughtful): • the feasibility of the test or intervention • "the test might work but if my practice can't afford to buy the machine it doesn't matter" • the patient or disease-oriented nature of the outcome • "did they measure children's ability to use improved function in play activities?" • the congruence between the operational definition and the research question e.g. "whether their method of measuring the outcome is a realistic representation of the outcome we care about"	9: **Less thoughtful discussion of the relevance of the independent and dependent variables used in the study. May include specific concepts or examples without clear rationale.** May refer to same items listed in 'excellent' but without demonstrating depth of understanding.	5: **Response implies consideration of how well the study addresses the question at hand, but offers little discussion about why this may be important** • e.g. "what are the variables?"; • "does it answer my question?"; • "the outcome measure"; • "the purpose of the study"; • "will it impact my practice?"; • "length of follow-up"	0: No discussion of the research question and variables used to answer it.
b: Description of Subjects	12: **Includes both:** A clear expression of the importance of the link between the study subjects and target population. AND At least one example of a relevant disease or demographic characteristic • e.g. "were the patients similar to mine in terms of age and race?" or • "was it a hospital or clinic sample like my patients?" or • "did patients have same level of disease	9: **Includes one but not both:** A clear expression of the importance of the link between the study subjects and target population **OR** At least one example of a relevant disease or demographic characteristic e.g. "is the patient like mine?" or "education level of population"	5: **Response implies consideration of the study subjects, but offers no discussion of the connection between study subjects and target population or specific characteristics of the sample** • e.g. "is it an appropriate sample?" or • "what was the response or participation rate?" or • "what were the exclusion	0: No discussion of the characteristics of the research subjects.

QUESTION #6: When you find a report of original research related to your clinical question or any others, what characteristics of the study will you consider to determine if its findings are valid? (You've already addressed relevance, and question 7 will ask how to determine the importance of the findings. For this question, please focus on the validity of the study.)					
	Excellent	Strong	Limited	Minimal	Not evident
a: Internal Validity	24: Lists or describes at least 5 issues important to internal validity, such as: • Appropriateness of study design • Adequacy of blinding • Allocation concealment • Randomization of group assignment • Invalid or biased measurement ("followed own protocol?") • Importance of comparison or control group • Intention to treat analysis • Consideration of appropriate covariates ("were other relevant factors considered?") • Conclusions consistent with evidence ("do the results make sense?") • Importance of follow-up of all study participants • Appropriate statistical analysis • Sample size / Power • Sponsorship • When study was conducted • Confirmation with other studies • Valid outcome measures	18: Identifies 3-4 specific issues as above.	10: Identifies 2 specific issues as above.	5: Mentions internal validity or lists one specific concept from examples above.	0: None of the above present

NOTE: RESPONSES TO QUESTIONS 5, 6, AND 7, CAN BE APPLIED TO ANY PORTION OF THE GRADING RUBRIC FOR THOSE ITEMS

QUESTION #7: When you find a report of original research which relates to your clinical question or any others, what characteristics of the findings will you consider to determine their magnitude and significance (clinical and statistical)?					
	Excellent	Strong	Limited	Minimal	Not evident
a: Magnitude	12: Response must clearly discuss both: • clinical significance ("what is the clinical significance?" or "how large a difference was found?", does change exceed MCID) AND • example(s) of effect size measurements (e.g., specificity, sensitivity, likelihood ratio of a test, number needed to treat, relative risk, absolute risk reduction, mean difference for continuous outcomes, positive or negative predictive value)	9: Response discusses one but not both: • clinical significance ("what is the clinical significance?" or "how large a difference was found?") OR • example(s) of effect size measurements (e.g., specificity, sensitivity, likelihood ratio of a test, number needed to treat, relative risk, absolute risk reduction, mean difference for continuous outcomes, positive or negative predictive value)	5: Response only suggests consideration of clinical significance or size of effect. • e.g. "does it matter?" "will it impact my practice" or • e.g. mentions "Minimal Clinically Important Difference" but does not explain how this value would be used to determine clinical significance)		0: None of the above present
b: Statistical Significance	12: Well-reasoned and thoughtful discussion of the indices of statistical significance, including at least 2 specific examples of important related concepts such as: • p-values • confidence intervals • power • precision of estimates • Type 1 or Type 2 error	9: Lists more than one concept (as above) with insufficient or absent discussion (e.g. "p-value and confidence intervals") OR Lists and discusses only one concept (e.g. "p-value less than <.05")	5: Mentions need to assess statistical significance or names only one concept from above without further discussion (e.g. "p-values", "statistically significant")		0: None of the above present

NOTE: RESPONSES TO QUESTIONS 5, 6, AND 7, CAN BE APPLIED TO ANY PORTION OF THE GRADING RUBRIC FOR THOSE ITEMS

QUESTION #8: For the clinical scenario that you chose, list up to two questions that you would ask the patient/family to gain a better understanding of his or her personal preferences and/or circumstances regarding your clinical question.					
	Excellent	Strong	Limited	Minimal	Not evident
a: Question 1	8: Question is likely to elicit important information about patient preferences, values, circumstances, expectations, and/or motivations that are will directly impact clinical care.		4: Question is general but addresses issues relevant to understanding the patient's perspective	2: Question is general and does not address issues specific to the patient's perspectives e.g. Standard question from subjective evaluation not specific to patient perspectives e.g. Yes/No or factual questions that are unlikely to elicit details about patient perspective	0: No question or not an actual question • "past medical history" or • "preferences"
b: Question 2	8: same as above but elicits different information than the first question (otherwise 0)		4: same as above but elicits different information than the first question (otherwise 0)	2: same as above but elicits different information than the first question (otherwise 0)	0: No question or not an actual question

Item #9 was dropped from the final version of the modified FT due to poor psychometric performance.

QUESTION #10:
A study of the diagnostic accuracy of exercise treadmill testing (ETT) in diagnosing of coronary artery disease (CAD) included 96 women with suspected CAD, 29 of whom were subsequently determined to have CAD (>50% stenosis in one or more coronary vessels). Of those with CAD, 9 had an abnormal ETT. Of the 67 patients determined not to have CAD, 32 had an abnormal ETT.

*** Alternative text using natural frequency values:**
A study of the diagnostic accuracy of exercise treadmill testing (ETT) in diagnosing of coronary artery disease (CAD) included 120 women with suspected CAD, 30 of whom were subsequently determined to have CAD (>50% stenosis in one or more coronary vessels). Of those with CAD, 10 had an abnormal ETT. Of the 90 patients determined not to have CAD, 30 had an abnormal ETT.

	Excellent	Strong	Limited	Minimal	Not evident
a: Based on these results, the sensitivity of ETT for CAD is…	4: 0.31; 31%; 9/29 *4: 0.33; 33%; 10/30	3: Within 5%: 26-36%			0: No answer or wrong answer
b: Based on these results, the positive predictive value of ETT for CAD is…	4: 0.22; 22%; 9/41 *4: 0.25; 25%; 10/40	3: Within 5%: 17-27%			0: No answer or wrong answer
c: Based on these results, the likelihood ratio positive for an abnormal ETT for CAD is…	4: 0.65; 31/47 *4: 1.0; 0.333/1-0.666	3: Within 5%: 0.60-0.70			0: No answer or wrong answer

Rounding is acceptable (*e.g.* 21.9 to 22 is acceptable)

QUESTION #11: A recent randomized trial of pregnant women with incontinence found that after pelvic floor training 20% had incontinence compared to 32% in a control group at 3 mos after delivery. Alpha level for the study was set at the 0.05 significance level. *** Alternative text using natural frequency values:** A recent randomized trial of pregnant women with incontinence found that after pelvic floor training 20% had incontinence compared to 30% in a control group at 3 mos after delivery. Alpha level for the study was set at the 0.05 significance level.				
a: The absolute risk reduction for recurrent events is…	4: 0.12; 12% *4: 0.10; 10%	3: within 2%: 10-14%		0: No answer or wrong answer
b: The relative risk reduction for recurrent events is…	4: 0.38; 38%; 12/32 *4: 0.33; 33%; 10/30	3: within 2%: 36-40%		0: No answer or wrong answer
c: The number needed to treat (NNT) to prevent one recurrent event is…	4: 9; 1/0.12 *4: 10; 1/0.10	3: within 1: 8-10		0: No answer or wrong answer
d: The p-value indicating statistically significant difference between the groups would be…	4: <0.05	3: 0.05		0: No answer or wrong answer

QUESTION #12: The same study described in question 11 revealed a relative risk of incontinence of 0.61 for the women receiving pelvic floor training. This suggests that pelvic floor training treatment reduces risk for incontinence. We wonder if this difference is statistically significant, so we look at the confidence interval. Give an example of a confidence interval that would support the conclusion that the rate of incontinence was indeed (statistically) different for these two treatment groups.

*** Alternative text using natural frequency values:**
The same study described in question 11 revealed a relative risk of incontinence of 0.66 for the women receiving pelvic floor training. This suggests that pelvic floor training treatment reduces risk for incontinence. We wonder if this difference is statistically significant, so we look at the confidence interval. Give an example of a confidence interval that would support the conclusion that the rate of incontinence was indeed (statistically) different for these two treatment groups.

	Excellent	Strong	Limited	Minimal	Not evident
a	4: Indication that any CI that does not include 1 would indicate statistical significance				0: Other

QUESTION #13: Which study design is best for a study about diagnosis?

a	4: Cohort Study; Cross Sectional Study; Comparison with gold standard; systematic review		0: Other

QUESTION #14: Which study design is best for a study about prognosis?

a	4: Cohort; prospective; longitudinal; systematic review		0: Other