A Statistical Analysis of Case Specificity and Intervening Variables Affecting Diagnostic performance

[1]Dr. Fadi Munshi, [2]Dr.Mohammed yunus
[1.] Assistant Professor, Medical Education Department, King Fahad Medical City, King Saud Bin Abdul-Aziz University, KSA.
[2.] Department of Mechanical Engineering, College of Engineering and Islamic Architecture, Umm Alqura University, Makkah, KSA.

ABSTRACT

In medical education, the generalizability of diagnostic performance and problem solving strategy within and across cases from two similar clinical presentations (CP) depends on many variables such as disease difficulty, case typicality, CP, and problem solving strategy have been described in the literature. In order to examine the effect and degree of impact of intervening variables on diagnostic performance, two types of data were analyzed, diagnostic performance scores obtained from written clinical vignettes and problem solving strategies abstracted from written think aloud exercises. Classical Test Theory and Item Response Theory were two measurement frameworks used to analyze data and triangulate study findings. Cross-tabulation and Chi-square were used to investigate the consistency in using problem solving strategies within and across cases from two similar CPs. Also, Multi-Facet Rasch Modelling (MFRM) was employed to determine degree of impact on diagnostic performance by the independent variables. This study answered three research questions related to diagnostic generalizability and intervening variables that contribute to diagnostic performance. Most of the diseases included in this study showed high correlation coefficients and high discriminating measures and estimated measures from two different measurement theories supported predictability of performance from one case to another. The research model reveals the factors that influence diagnostic performance are disease difficulty and type of CP. This research provides evidence that the diagnostic strategy utilized can alter diagnostic performance. Further support was found for schema-based instruction in medical schools to enhance student diagnostic performance. These study findings set the base for future research.

Key words: Clinical presentations, Chi-Square, Classical Test Theory, Diagnostic performance, Disease difficulty, FACETS software, Item response theory, Rasch model.

1. INTRODUCTION

Errors in the healthcare are causing death and injury (Kohn, Corrigan, & Donaldson, 2000). The errors which resulted in medical injury can be classified as diagnostic, treatment, preventive and communication failure (Leape, 1994). Diagnostic errors are categorized into errors or delay in diagnosis, failure to employ indicated tests, use of outmoded tests of therapy, and failure to act on results of monitoring or testing. Physicians must demonstrate medical knowledge, patient care, communication, practice-based learning, professionalism, and system-based practice competencies in their attainment of competency in diagnosis. Physician's ability in diagnostic performance is determined by how a disease presents in terms of various combinations of signs and symptoms using a number of different case presentations for a specific disease, as success in diagnosing a clinical case is content/knowledge dependent and process dependent (Harasym et al., 2008; Papa, Oglesby, Aldrich, Schaller, & Cipher, 2007).

However, it is unknown whether the generalization in diagnostic performance will occur within a CP – that is, the predictability of performance above levels, across levels, and between CPs. Therefore, there is a need to investigate if generalization exists. With the content and case specificity in mind, we chose to study two similar CPs. The purpose of this research was: (1) to investigate the degree to which diagnostic performances will generalize within a similar cluster of diseases (i.e., CP) and across a dissimilar cluster and (2) to identify the factors that are anticipated to influence diagnostic performance generalization.

1.1. Null Hypotheses of Data Analyses

The following null hypotheses were assumed:
1-Diagnostic accuracy will not generalize within CP cases and across two similar CPs;
2-Problem solving strategies utilized will not be consistent from one case to another and across CPs;
3-There will be no difference in diagnostic performance by students using different diagnostic strategies.

1.2. Research summary

The work carried out can be divided into three research questions, variables involved in it and type of statistical analysis is used as shown in table 1.

Table 1. Method Summary

Research Question	Variables	Statistical Analyses
What is the underlying structure within clinical cases, and across two similar clinical presentations?	Total score from 4 clinical vignettes on each of the 19 diseases.	Principal Component Analysis
How consistent are diagnostic strategies across two cases in the same and across two similar CPs?	Categorical Variable: Clinical problem solving approaches on each CP.	Cross-Tabulation Chi Square
What is the effect of disease difficulty, case typicality, CP, and clinical problem solving strategy on diagnostic performance?	Dependant: Score on each clinical (total 76). Independent: disease, typicality, type of CP, clinical problem solving strategy for each CP.	Multi Facets Rasch Measurement (MFRM)

1.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) and Factor Analysis (FA) are two variable reduction strategies used to examine the relationship between and within CP/disease variables. Although, some researchers use PCA and FA interchangeably, they differ conceptually. As stated in O'Rourke, Hatcher, and Stepanski (2005, p 436). Both (PCA & FA) are methods that can be used to identify groups of observed variables that tend to hang together empirically.

Both procedures generally tend to provide similar results. PCA reduces the number of observed variables to fewer principal *components,* which account for a *maximum amount of variance* of the observed variables. In the present study, PCA was employed as a variable reduction strategy to explain the variance and identify the groups of inter-related items or components.

1.4. Cross Tabulation Table (CTT) and Chi-Square Methods

After determining the problem solving strategies used by each participant for chest discomfort and dyspnea, a cross-tabulation between the clinical problem solving categories was conducted. Cross-tabulation is a method used to analyze categorical variables.To examine if the two problem solving strategies used in chest discomfort and dyspnea CPs are independent, Chi-square was computed. The Chi-square statistic is the primary statistic used for testing the statistical significance of the cross-tabulation table. If the variables are dependant, then the results of the statistical test are significant and the null hypothesis is rejected. This means that there is a relationship between the variables (Field, 2000).

1.5. Multi-Facet Rasch Modelling (MFRM)

It is statistical analysis method that is based on Item Response Theory (IRT). MFRM generate a composite score based on the contribution of multiple facets and provides the percentage of variance removed from the true score (Bond & Fox, 2001). MFRM was computed to determine the degree of impact (i.e., variance accounted for) by each of the independent variables. The assumption of uni-dimensionality for this statistical procedure was determined by PCA and fit statistics. The data were analyzed by FACETS (version 3.71.0) (Linacre, 2013). The program used the scores of diagnostic performance to estimate the effect of disease difficulty, case typicality, and problem solving strategy on examinee abilities. FACETS calibrated onto the same equal-interval scale (Linacre, 2005). This generates a logit-based scale for interpreting the results of the analyses. IRT is a probabilistic psychometric measurement model used to estimate ability and item characteristics on the trait measured (Bond & Fox, 2001). IRT presumes some assumptions that when understood help explain the theory (Bechger, Maris, Verstralen, & Berguin, 2003; Morris et al., 2006; Schaefer, 2008). These assumptions are

a. One common factor accounts for all item co-variances and relations between them.
b. The observed response take a particular form called an item characteristic curve.
c. All items must contribute in a meaningful way to the attribute being investigated.

There are three main approaches in assessing the dimensionality of an examination: 1) using Principal Component Analysis (PCA) only, 2) using fit statistics only, and 3) using fit statistics, then exposing the residuals to a PCA (Tennant & Pallant, 2006). For a 1-parameter IRT model, which includes difficulty only, a sample size of 30 for robust decision making is required (Linacre, 1994). Rasch modelling is considered a 1-parameter IRT model. In the current study, a form of Rasch modelling, MFRM, is computed therefore the sample size of 175 participants was deemed sufficient.

2. METHODOLOGY

The present study uses PCA and FA, CTT and Chi-Square and MFRM to detect the degree of impact on diagnostic performance by the independent variables such as disease difficulty, case typicality, type of Clinical Presentations (CP) and problem solving strategy. It incorporates 175 candidates, 76 disease items, four typicality levels, two CPs, and three problem solving strategies.

Two types of data were collected; diagnostic performance from clinical vignettes and abstracted clinical problem solving strategies from written think aloud (WTA) exercises. All participants from the four universities were invited to voluntarily participate in a 5 hour session in which medical clerks responded to 76 clinical vignettes and 4 WTA exercises. In return, each participant received feedback on diagnostic performance categorized by disease and CP. Chest Discomfort and Dyspnea schemes were used to investigate diagnostic performance generalizability within and across CPs. Table 2 below shows the cardiac causes of chest discomfort.

Table 2. Diseases Included in the Study sorted by Clinical Presentation

Dyspnea	Chest Discomfort
Acute Respiratory Distress Syndrome	Acute Coronary Syndrome
Anemia	Anxiety/Panic disorder
Asthma	Constrictive Pericarditis
Cardiac Tamponade	Lung Cancer
Chronic Obstructive Pulmonary Disease	Peptic Ulcer Disease
Congestive Heart Failure	Pneumonia
Pulmonary Embolism	Pneumothorax
Pulmonary Hypertension	Stable Angina
Sarcoidosis	Valvular Regurgitation
	Valvular Stenosis

2.1. Methodology for PCA

The first step in the analyses involved conducting two separate PCAs for the nine cases in chest discomfort and ten cases of dyspnea. The purpose was to observe if the items within each CP group together in the same manner as they are grouped in the CP schemes – that is, by etiology, body system, or CP.

The second step was to determine if the 19 diseases cluster together into two CPs or sub-categories (i.e., body system or etiology), or if they cluster together under one group as a general diagnostic performance construct.

2.2. CTT and Chi-Square Methodology

The association between two categorical variables has been established in this paper. Statistics commonly used for assessing the statistical significance and strength of association of crosstabulated variables were explored. The statistical significance of the observed association is commonly measured by the chi-square statistics (χ 2). The strength of the association, or degree of the association, is important from a practical or substantive perspective. Generally, the strength of association is of interest only if the association is statistically significant.

2.3. MFRM Methodology

MFRM was computed to determine the degree of impact (i.e., variance accounted for) by each of the independent variables. The assumption of uni-dimensionality for this statistical procedure was determined by PCA and fit statistics. The data were analyzed by FACETS (version 3.71.0) (Linacre, 2013). The program used the scores of diagnostic performance to estimate the effect disease difficulty, case typicality, and problem solving strategy on examinee abilities. FACETS calibrated diagnostic performance scores, disease difficulty, CP, case typicality, and problem solving strategies onto the same equal-interval scale (Linacre, 2005). This generated a logit-based scale for interpreting the results of the analyses.

3. RESULTS & DISCUSSION

3.1. Principal Component Analysis and Factor Analysis

This is the answer for first research question. Across two similar CPs, low to moderate correlations between cases were observed. A forced two-factor PCA was conducted. The loading of 14 out of 19 diseases on component one resulted in a new dimension similar to the original two CPs. These 14 diseases were a mix of from both CPs and from varying levels of the scheme. Referring to the original CPs, this suggests that most of the diseases did not have a distinctive association with the one CP. Rather, they may be associated with other factors, such as a more general construct of diagnostic accuracy, a compound anatomical or pathophysiologic construct, a technologically-based construct of relatively defining laboratory findings.

Reece et al. (2008) studied OSCE performance scores. Using confirmatory FA, the authors identified best-fit indices with a 3-component model. The authors argue that these three components are history taking, physical examination, and communication skills. Perhaps skills tend to distinct attributes. However, it is possible that diagnostic accuracy may be dependent on several inter-related variables such as disease difficulty, similarity of CPs, and diagnostic strategy used.

3.2. Results of CTT and Chi square

This provides answer to second question of Research. The results estimated that generalizable skills and case specific knowledge each accounted for 20% of the variance in performance scores. This study found that case difficulty on two similar CPs accounted for the largest amount of variance (29.71%). This amount is greater than that reported by Wimmers and Fung.This study supports the findings of similar studies of problem solving strategies and diagnostic performance (Coderre et al., 2003; McLaughlin et al., 2007). ANOVA indicated that participants using a Scheme Inductive problem-solving strategy showed evidence of better diagnostic performance. Thus, students who used scheme induction on average scored 44.7%

higher than students who used guessing; 15.5% higher than those who used hypothetical deduction (HD). More importantly, is that these students were all taught HD reasoning throughout their 6 years of medical school. In other words, advanced students will eventually gravitate to SI reasoning in spite of the predominant teaching of HD reasoning. It appears to be a natural outcome.

Table 2.a. CTT of Problem Solving Strategies used for Chest Discomfort and Dyspnea

Problem Solving Strategy Dyspnea (DYS)						
			Guessing DYS	Hypothetical Deductive DYS	Scheme Inductive DYS	Total
Problem Solving Strategy Discomfort (CD Chest)	Guessing CD	Count	13	10	0	23
		% within Approach CD	**56.5%**	43.5%	0.0%	100.0%
	Hypothetical Deductive CD	Count	11	105	10	126
		% within Approach CD	8.7%	**83.3%**	7.9%	100.0%
		Count	0	13	13	26
	Scheme Inductive CD	% within Approach CD	0.0%	50.0%	**50.0%**	100.0%
Total		Count	24	128	23	175
		% within Approach CD	13.7%	73.1%	13.1%	100.0%

3.3. Results of MFRM

To study the effect of disease difficulty, case typicality, type of CP, and problem solving strategy on diagnostic performance - that is, candidate ability, MFRM was conducted using FACETS software. It provides answer to third research question. MFRM was computed with 175 candidates, 76 disease items, four typicality levels, two CPs, and three problem solving strategies. The variance explained by the Rasch measures was only 24.15%. This variance-explained is dependent on the variances by the measures of the elements in the facets.

All the facets have variances smaller than one logit because the standard deviations were low. Therefore, variance explained by Rasch measures can only be small (Linacre, 2005). This finding of low standard deviations was attributed to the fact that the 76 cases were assigned scores ranging from zero to one. To adjust for this low variance, the observed scores were re-grouped. Performance scores on the four clinical vignettes that represent each disease were summed. This led to observed scores ranging from zero to four for each of the 19 diseases. In addition, the scores for typicality and diagnostic strategy were summed to reflect a total score for each of the 19 diseases.

3.3.1. Model Fit and Uni-Dimensionality

MFRM analysis was run and nine disjoint subsets in the CP facet were detected and found when the elements of the facets are not sufficiently crossed. In the current data, to resolve diseases are nested within case typicality and CP were converted into “demographic" facets and are not included in the measurement model. After the adjustment of the model, subset connection in the data was achieved with the measurement model including candidates, disease, and problem solving approach. Thus, the estimates of typicality and CP were excluded from the analysis.

Table3. Forced 2-Factor Solution Principal Component Analysis

Item	Disease	Communality	Rotated Components	
			CP	Difficult Diseases1
Pepulc	Peptic ulcer	0.58	0.76	0.02
Anx	Anxiety	0.37	0.54	0.27
Chf	Chronic heart failure	0.48	0.64	0.27
Valvreg	Valvular regurgitation	0.38	0.36	0.50
Acs	Acute coronary syndrome	0.58	0.75	0.13
Stabangi	Stable angina	0.35	0.56	0.17
Pneumonia	Pneumonia	0.61	0.75	0.20
Pneumot harax	Pneumothorax	0.57	0.52	0.54
Constperi	Constrictive pericarditis	0.44	0.16	0.64
copd	chronic obstructive pulmonary disease	0.42	0.59	0.26
pulmembo	pulmonary embolism	0.67	0.77	0.27
anemia	Anemia	0.41	0.51	0.38
sarcoid	Sarcoidosis	0.51	0.59	0.39
lungcan	lung cancer	0.55	0.73	0.08
cardtamp	cardiac tamponade	0.42	0.01	0.65
ards	acute respiratory distress syndrome	0.43	0.18	0.62
pulmhyp	pulmonary hypertension	0.46	0.23	0.64
valvsten	valvular stenosis	0.63	0.74	0.14
asthma	Asthma	0.18	0.39	0.18
# of Diseases	19	Cross Loading	13	4
Cronbach's α	0.91			2
Kaiser-Meyer-Olkin (KMO)	0.92			
Bartlett's test of Sphericity	0.000		32.84%	14.88%
Total variance explained	47.71%	Eigenvalue	6.24	2.83

*Diseases1 = 4 difficult diseases clustered together likely due to reduced variance in clerks' diagnostic scores.

Upper Case: Chest Discomfort Cases Lower Case: Dyspnea Cases

Model fit and uni-dimensionality were estimated using standard residuals and infit-outfit indices. Standardized residuals mean was 0.01 and the sample standard deviation was 1.00. These results indicate that the data fit the Rasch model.

Uni-dimensionality of data is a MFRM assumption. Fit statistics revealed four candidates (2.28%) out of 175 with infit Z standardized statistics (Zstd) that were less than the critical value of -2 and ten (5.71%) with outfit Zstd statistics that were greater than the critical criteria of +2.0 (Bond & Fox, 2001). Of the 19 disease items, one (5.26%) had infit Zstd statistics less than -2 and one (5.26%) had Zstd statistics greater than the criterion of +2.0. The percentage of candidates and diseases showing acceptable fit statistics did not fall below 90%. These findings with the PCA conducted in research question one support the uni-dimensionality of the data as shown in table 3.

3.3.2. Effect of Facets

Reliability estimates were estimated for each facet and the gold standard was set at 0.80. The estimated reliabilities for the facets candidates, diseases, and problem solving approach are 0.86; 0.99; and 0.99, respectively. This indicates high reliability.
MFRM analysis revealed that 56.80% of the main effects variance was explained by Rasch-measures. Of the total systematic variance, candidate ability accounted for 17.74.8%, differences in disease difficulty resulted in 29.71% variance, and 9.34% was due to problem solving strategy.

3.3.3. Candidate Ability

The abilities of 175 clinical clerks ranged from +192 till -131 logits. According to Rasch estimated true diagnostic ability, candidate 102 had the highest ability of +192 logits. The raw score of this candidate was 68 out of 76. In contrast, candidate 111 had a raw score 15 out of 76, and had an ability Rasch true measure of -131.

3.3.4. Difficulty and Discrimination Indices of Diseases

Disease difficulties are listed in Appendix A. Diseases varied in difficulty with a range of +148 to -91 logits. The most difficult disease to diagnose was acute respiratory distress (total score 202 and logit score of +147) while the easiest was peptic ulcer disease (total score 589 and logit score of -91). Without altering other estimates, an estimate of item discrimination is computed in MFRM (Linacre, 2013). Disease discrimination indices for 14 out of 19 items were equal or close to the desired Rasch expectation of $di = 1$.

3.3.5. Clinical Problem Solving Approach

Clinical problem solving approach MFRM measurements are reported. SI reasoning was the most discriminating with a of $di = 1.14$. Clerks who used guessing had a total score of 637 (+56 logits), those who used HD reasoning had a total score of 6288 (-1 84 logits), and students who used scheme induction had a total score of 1492 (-34 logits). This finding indicates that clerks who tended to use guessing as a diagnostic strategy were least likely to obtain the correct diagnosis (mean score of 26.64), clerks who tended to use HD reasoning were more likely to get the correct diagnosis (mean score of 50.31), and clerks who used scheme induction were most likely to answer correctly (mean score of 62.38). In both CPs, there was a statistically significant difference among the three strategy groups as determined by one-way ANOVA, $F(2,172) = 85.19$, $p < 0.001$ for chest discomfort cases and, $F(2, 172) = 84.87$, $p < 0.001$ for dyspnea cases.

3.4. Contributions to Medical Education Literature

This study debates the "all or none" notion of case specificity. The findings of this study indicate that a degree of diagnostic performance predictability exists in the context of cases that are conceptually linked together in a CP and across two similar CPs.

3.5. Implications for Testing Agencies

Testing agencies can benefit from the findings of this research in reducing the number of items needed to evaluate competency. In assessment, greater generalizability means fewer items required to test competency. In our current study we found a high degree of predictability of diagnostic performance between cases within and across two similar CPs Using the Spearman-Brown prophecy formula, which is a formula used to predict the reliability of a test after changing the test length, the items can be reduced. In this study, as shown is table 4, results showed that it would take around 20-40 cases within a CP to reliably estimate diagnostic performance. If there is a limited amount of testing time, the measure is uni-dimensional, diagnostic performance is analyzed using MFRM, and there is a reduced need to ensure a representative sampling of items within a CP given the generalizability of performance from one case to another, the reduction in cases may be possible This proposition is theoretical and would require further investigation.

Table 4: Number of items required for each CP to obtain an estimated 0.80 reliability.

Study	CP	# of Items	Reliability	Target Reliability	Spearman Brown # of Items
Current Finding CTT	Chest Pain	36	0.88	0.80	20
Current Finding CTT	Dyspnea	40	0.84	0.80	30
Current Finding MFRM	Chest Pain	36	0.81	-	36*
Current Finding MFRM	Dyspnea	40	0.83	0.80	33

*Not computed by Spearman-Brown prediction formula.

4. CONCLUSION

This study answered three research questions concerning diagnostic generalizability and intervening variables that contribute to diagnostic accuracy. Most of the diseases chosen in this study showed high correlation coefficients and high discriminating measures. These estimated measures from two different measurement theories support predictability of performance from one case to another within two similar CPs.

The research model illuminates some factors that influence diagnostic performance. Disease difficulty and diagnostic strategy were two variables found to impact diagnostic performance. Further research is required to identify other influencing variables.

This study also provides evidence that the direction of clinical reasoning utilized enhances diagnostic performance. Schema-based instruction is a recommendation for medical schools to apply as the empirical evidence provided in this study linked better diagnostic accuracy with forward reasoning. Finally, study findings showed consistent use of the same problem solving strategy within and across two similar CP cases. Generalizability of diagnostic performance and problem solving strategy within and across cases from two CPs are main contributions to the medical education literature that provide ample opportunities for further research.

REFERENCES

1. Bechger, T., Maris, G., Verstralen, H., & Beguin, A. (2003). *Using Classical Test Theory in combination with Item Response Theory.* Applied Psychological Measurement, *27,* 319-334.

2. Bond, T. & Fox, C. (2001). *Applying the rasch model: fundamental measurement in the human sciences*. Psychology Press.

3. Bordage, G. (2007). Prototypes and semantic qualifiers: From past to present. *Medical Education, 41,* 1117-1121.

4. Chi, M. T. H. (2006).Two approaches to the study of experts' characteristics. In K.A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), The Cambridge handbook of expertise and expert performance (pp. 21-30). New York, NY, US: Cambridge, University Press.

5. Harasym, P. H., Tsai, T. C., & Hemmati, P. (2008). *Current trends in developing medical students' critical thinking abilities*. The Kaohsiung journal of medical sciences, 24, 341-355. Hardie, E. M. (2008). Current methods in use for assessing clinical competencies: What works? *Journal of Veterinary Medical Education, 35,* 359-368.

6. Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. D. (2000). *To err is human: building a safer health system*. Washington, D.C.: National Academu Press. Kalyuga, S. & Hanham, J. (2011). Instructing in generalized knowledge structures to develop flexible problem solving skills. *Computers in Human Behavior, 27,* 63-68.

7. Leape, L. L. (1994). *Error in medicine*. JAMA: The Journal of the American Medical Association, 21, 1851-1857.

8. Linacre, J. M. (1994). *Sample size and item calibration stability*. Rasch Measurement Transactions, 7, 328.

9. Linacre, J. (2005). *Rasch dichotomous model vs. one-parameter logistic model*. Rasch Measurement Transactions*, 19,* 1032. 108.

10. Linacre, J. M. (2013). *A user`s guide to FACETS*. Winsteps.

11. Morris, G., Martin, L., Harshman, N., Baker, S., Mazur, E., Dutta, S. (2006). *Testing the test: item response curves and test quality*. American Journal of Physics*, 74,* 449-453.Norman, G. R. (2008). The glass is a little full of something: revisiting the issue of content specificity of problem solving. Medical Education, 42, 549-551.

12. Papa, F., Oglesby, M., Aldrich, D., Schaller, F., & Cipher, D. (2007). *Improving diagnostic capabilities of medical students via application of cognitive sciences-derived learning principles. Medical Education, 41,* 419-425.

13. Schaefer, E. (2008). *Rater bias patterns in an EFL writing assessment*. Language Testing, *25,* 465-493.

14. Tennant A. & Pallant J.F. (2006). *Uni-dimensionality matter! (a tale of two smiths?).* Rasch Measurement Transactions, 20, 1048-1051.

Appendix A: MFRM Disease Measurement Report

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis	Nu Disease
202	175	1.15	1.02	147	7	1.19	1.7	1.16	1.4	.79	.28	10 Acute Respiratory Distress Syndrome
254	175	1.45	1.33	115	6	1.06	.6	1.05	.5	.91	.30	12 Pulmonary Hypertension
304	175	1.74	1.62	88	6	1.03	.2	1.10	.9	.88	.25	1 Constrictive Pericarditis
311	175	1.78	1.66	84	6	1.23	2.1	1.32	2.8	.58	.20	6 Cardiac Tamponade
315	175	1.80	1.68	82	6	1.03	.3	1.00	.0	.99	.39	13 Valvular Regurgitation
396	175	2.26	2.19	38	6	.54	-5.5	.56	-5.3	1.44	.39	11 Congestive Heart Failure
427	175	2.44	2.41	19	6	1.06	.6	1.15	1.4	.76	.27	3 Asthma
448	175	2.56	2.53	8	6	1.08	.8	1.03	.3	1.00	.44	14 Valvular Stenosis
467	175	2.67	2.66	-3	6	1.13	1.2	1.08	.7	.84	.34	16 Stable Angina
481	175	2.75	2.76	-12	7	1.09	.9	1.06	.6	.94	.37	4 COPD
505	175	2.89	2.92	-27	7	1.15	1.4	1.08	.7	.98	.39	17 Sarcoidosis
508	175	2.90	2.93	-28	7	1.02	.2	.96	-.3	1.09	.40	19 Pneumothorax
508	175	2.90	2.94	-29	7	.85	-1.4	.80	-1.9	1.28	.45	5 Pulmonary Embolism
521	175	2.98	3.02	-37	7	.93	-.6	.92	-.7	1.10	.44	15 Acute Coronary Syndrome
527	175	3.01	3.07	-42	7	.94	-.5	.91	-.8	1.08	.39	7 Anemia
530	175	3.03	3.09	-44	7	.93	-.5	.91	-.8	1.09	.36	2 Lung Cancer
557	175	3.18	3.26	-64	7	1.03	.2	1.05	.4	1.00	.33	9 Anxiety/Panic disorder
567	175	3.24	3.32	-72	8	.92	-.6	.88	-.9	1.15	.41	18 Pneumonia
589	175	3.37	3.46	-91	8	.90	-.8	.82	-1.2	1.10	.37	8 Peptic Ulcer Disease
443.0	175.0	2.53	2.52	7	7	1.01	.0	.99	-.1		.36	Mean (Count: 19)
111.0	.0	.63	.71	66	0	.15	1.6	.16	1.7		.07	S.D. (Population)
114.0	.0	.65	.73	68	1	.15	1.6	.17	1.7		.07	S.D. (Sample)

Appendix B: MFRM Clinical Problem Solving Approach Measurement Report

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis	Exact Agree. Obs %	Exact Agree. Exp %	N Clinical Problem Solving Approach
637	447	1.43	1.98	56	4	1.02	.3	1.05	.7	.94	.22	.0	.0	1 Guessing
6288	2414	2.60	2.63	-1	2	1.03	.9	1.01	.3	.98	.32	.0	.0	2 Hypothetical Deductive
1492	464	3.22	2.99	-34	5	.88	-1.6	.83	-2.2	1.14	.26	.0	.0	3 Scheme Inductive
2805.7	1108.3	2.42	2.54	7	4	.98	-.1	.97	-.4		.27			Mean (Count: 3)
2487.0	923.3	.74	.42	37	1	.07	1.1	.09	1.3		.04			S.D. (Population)
3045.9	1130.8	.91	.52	46	2	.08	1.4	.12	1.6		.05			S.D. (Sample)