Evaluating the Construct Validity of Basic Science Curriculum Assessment Instrument for Critical Thinking: A Case-Study

Chau-Kuang CHEN
Office of Institutional Research, Meharry Medical College
Nashville, TN 37208, USA

Adriana Marie HORNER
Medical Student, Meharry Medical College
Nashville, TN 37208, USA

Michelle SCOTT
Medical Student, Meharry Medical College
Nashville, TN 37208, USA

and

Stephanie C. MCCLURE
Office of Student and Academic Affairs, Meharry Medical College
Nashville, TN 37208, USA

ABSTRACT

The Rasch model is a practical framework for evaluating a construct validity of assessment instruments. It is capable of determining how the measurement of person's ability (endorsement) and item difficulty matches with each other. This study aimed at evaluating the psychometric properties (reliability, validity, and utility) of a basic science curriculum assessment instrument. Special emphasis was placed on finding the strengths and challenges in the curriculum, and detecting the existence of multidimensional structure. A total of 130 medical students in academic year 2016/17 completed a 22-item assessment instrument. Three major steps were involved in this study. First, the parameters of person's ability and item difficulty were separately estimated. Second, infit/outfit mean square residuals and standardized residual variance from principal component analysis (PCA) were used to validate the unidimentionality assumption. Lastly, differential item functioning (DIF) was assessed to determine the fairness of the assessment instrument. As a result, the baseline measures of the strengths and challenges in medical curriculum were established for continuous quality improvement. However, the unexplained variance for the first contrast value of 3.08 in PCA was greater than the criterion of 2.0, which shows some degree of violation of the unidimensionality assumption. Therefore, this instrument must be further revised for future application.

Keywords: Construct Validity, Curriculum Assessment, Differential Item Functioning, Item Difficulty, Rasch Model

1. INTRODUCTION

Curriculum assessment provides a strong rationale for securing curriculum changes and helps faculty and administrators make effective decisions about program content [1]. It is a process of collecting and analyzing data from multisource assessment to promote student learning [2]. Additionally, it is a way of accomplishing accountability reporting purposes to meet an institution's accreditation mandates. Ultimately, a workable assessment model should help students identify and respond to their own learning needs.

Medicine is a discipline that brings basic scientific knowledge full circle; from basic science course teaching and learning, through applied clinical practice. To succeed in medical school, it is important for medical students to understand different facets of human disease processes such as treatment, drug indications and side effects, effects on other organ systems, and disease etiology and prevention [1]. Medical school curriculum must be well-coordinated and integrated to prepare future physicians for clinical problem-solving that requires the coordination and integration of basic scientific knowledge.

The U.S. Medical Licensing Examination (USMLE) scores are used by most residency program directors to assess students' capacity to potentially participate in their programs. The School of Medicine at the College chose to make subject board examinations and the comprehensive basic science exam mandatory for all medical students to better prepare them for the USMLE Step 1. In the past, academic promotion in the first two years was contingent upon passing the corresponding NBME subject board examination [3], although it was not part of the course grade. Starting in academic year 2016-2017, the score is designated a percentage of the grade (10-25%) within each course within the basic science curriculum, also known as the preclinical curriculum (a hybrid integrated curriculum). Thus, it is crucial for the basic science curriculum to offer adequate scientific knowledge throughout the preclinical years with students being assessed both by internal discipline examinations and also external standardized tests such as NBME subject boards, and the USMLE Step 1. As of 2017-2018, the NBME Customized Assessment Services is being utilized on a pilot basis to develop some examinations at midterm and for all end of course exams for the preclinical years. Utilization of this service might have its limitations due to few questions being available in the pool for some topics.

To fulfill its mission in providing excellent health science education for its students, the College has implemented a policy of curriculum assessment. This policy is an integral part of the process to monitor and improve the curriculum and quality of instruction. The student evaluations of the medical curriculum included clarity of objectives, organization of the course content, and contribution of the curriculum to student professional development. Student perceptions of the overall strengths and weaknesses of the medical curriculum, as well as suggestions for improvements, were solicited.

Critical Thinking in the Medical Curriculum: Definitions of critical thinking cover a broad range of concepts, depending on the profession, focus area, and knowledge base of the institution or person defining it. In the medical field, critical thinking skills are applied both in the context of didactic learning (students) and application of clinical skills (physicians). The USMLE study outline for board preparation includes comprehensive foci related to these abilities, such as clinical decision making, use of Evidence-Based Medicine (EBM) in practice, constructing individualized patient risk profiles, and cost/benefit analyses of treatments [4]. Induction, deduction, inference, analysis, evaluation, self-regulation, open-mindedness, and systematicity are often named and tested as reliable measures of critical thinking [5, 6].

It is important for faculty and administrators to view critical thinking and clinical reasoning as both abilities and dispositions of learners. Awareness that these skills can be taught, and that students may need motivation to gain them, is essential for improving these proficiencies. Professors and clinical preceptors must probe students to ask questions, discuss cases with peers, challenge assumptions, and justify reasoning for didactic and clinical decisions [7].

The knowledge and application of medicine is constantly changing, as does the pool of applicants who enter medical school each year. Students of the millennial generation are less likely to prefer "traditional" lecture format, passive instruction, and have less motivation to complete assignments that lack transparent relevance to their learning. These students, however, are also more adept at multitasking with media simulations and technology, thrive in peer group activities, and learn readily by performing tasks [6].

U.S. students also have diverse experiences in learning STEM content before medical school commencement, due to variation in teaching methods, curricula, and emphasis on retention of facts versus gaining conceptual knowledge. Current instructional methods in the STEM fields may prioritize factual recall over deeper understanding of scientific principles and utilization of problem-solving skills [8]. This may explain the finding that some characteristics of items in scientific reasoning assessments, including the presence of abstract concepts, specialist terms, and formulas, may inherently be more difficult for this generation of student learners [9]. The application of standardized and reliable assessments of critical thinking throughout the medical curriculum may assist in our goal to identify specific deficits and improve them.

Assessments of Critical Thinking: Over the past decade, growing interest in students' critical thinking abilities has spawned an increasing utilization of external measures of student learning outcomes by higher education institutions. Individuals scoring higher critical thinking scores on these exams later report less negative life events and better passing rates on standardized professional tests [10]. In contrast, lower critical thinking scores have been associated with lifestyle and professional difficulties, such as increased unemployment rates and credit card debts post-college graduation [11]. A variety of assessments of critical thinking, clinical judgment, and scientific reasoning are being used in the health professions, most notably the California Critical Thinking Dispositions Inventory (CCTDI, California Academic Press) and the Health Sciences Reasoning Test (HSRT, California Academic Press).

The lack of standardization across institutions, reliability issues, and costs of test administration have some critics questioning the functional utility of these exams. For instance, Performance-Based Assessments (PBAs) have shown wavering efficiency and reliability, possibly due to variation in scoring methods and other measurement errors unrelated to the characteristics of the cases. PBAs have failed to show the ability to predict which students later become more skilled diagnosticians [12]. On the other hand, assessments which closely mirror common student studying techniques, items used in board preparation questions banks, and schematic methods used by clinicians to reach diagnoses, have shown more success at identifying differences between novice and skilled learners. Some of these efficacious assessments incorporate script concordance testing; this method tasks learners with evaluation of relationships between components of a clinical problem, and estimation of changes in probability of a diagnosis when new findings are introduced to the scenario [12, 13]. Student improvement over time has been seen with tests of diagnostic pattern recognition (DPR) of patient signs and symptoms, and clinical data interpretation (CDI) of the impact of data changes on the probability that a diagnostic hypothesis is correct [14]. These exams share a common theme, in that they assess the skills most commonly used by medical students and clinicians in their daily work.

The HSRT is administered at the College to evaluate student critical thinking skills. Interestingly, a number of studies have not discovered significant leaps in HSRT scores between preclinical and clinical years of the curriculum [11, 15]. Other studies have shown significant differences in subscores in deduction and analysis categories between students and experts, which may be potential targets for improvement [15]. The College has used the HSRT for objective analysis of critical thinking in its incoming students since 2009. Students are shown their subscale scores and advised about how to strengthen areas needing improvement. As part of the institution's Quality Enhancement Plan (QEP) for 2017, the College wishes to improve the student HSRT scores, particularly in the area of inference [16].

The Item Response Theory (IRT) and Rasch Model can be used to examine the characteristics of student survey item, as well person responses to each item. For instance, person parameters can include measures of an individual's ability to endorse a particular item, or their strength of attitude. Item parameters can distinguish between characteristics such as item difficulty, discrimination, and guessing. Thus, an objective measure of a person's ability can be estimated regardless of the item characteristics (such as the difficulty of the question) [17]. The College is positioning itself for major revision of the medical school curriculum starting academic year 2018-2019. To identify particular areas that need improvement in the medical curriculum, the Rasch Model was applied to medical student surveys distributed throughout the current curriculum.

Rasch Model for Ordered Response Categories: The person-item map in the Rasch analysis matches person's ability distribution with item difficulty distribution side by side on a linear logit scale [18, 19, 20, 21, 22]. On the person-item map, the character "M" (mean) on the right side of item histogram has a logit value of zero that is equivalent to the odds of 1 and the probability value of .5, which splits all items into half with relatively difficult items on the top of the map compared to its counterpart of the relatively easy item on the bottom of the map. A linear logit line with characters "M" (mean), "S" (one standard deviation), and "T" (two standard deviations) on each side separates the person histogram on the left and the item histogram on the right. Persons approaching the top of the map demonstrate the most ability "endorsement" while persons close to the bottom show the least ability. Items nearing the top of the map display the most difficult items while items close to the bottom exhibit the least difficult items [20].

The primary assumption of the Rasch model should be unidimensional, which can be assessed based on residual-based principal components analysis (PCA). The amount of variance explained by total measures of persons and items in PCA should be at least 60% accounted for by the Rasch dimension [23]. The unidimensionality assumption can also be verified by examining item fit statistics. Infit and outfit mean square fit statistics provide summaries of the residuals (differences between actual responses and their estimated responses for individual items and persons). High item mean square fit statistics show a large number of unexpected responses that may be due to ambiguous wording, misleading statements, or not measuring the same construct, which lead to a violation of unidimensionality assumption. Some researchers set the acceptable values of infit and outfit statistics from 0.6 to 1.4 [24, 25]. In general, items with infit and outfit mean squares between 1.5 and 2.0 are considered to be unproductive. Items with infit and outfit statistics greater than 2.0 are treated as degrading measurements, which need to be removed [25].

Separation index, the spread of items or persons in standard units, should be at least 2 units [23]. Low item separation (item separation < 3, item reliability < 0.9) implies that the sample of persons is not large enough to confirm the item difficulty hierarchy of the assessment instrument. [22, 26]. Also, low person separation (person separation < 2, person reliability < 0.8) implies that the instrument may not be sensitive enough to distinguish between high and low performers. In this case, more items may be desired [22, 26].

Person ability and item difficulty estimates should be evenly distributed on linear interval scale if the items accurately target person's ability. Also, item estimates should remain constant across different person's characteristics rather than display significant difference in differential item functioning (DIF) if the invariance property is held true [27, 28, 29, 30]. In other words, a decent item should not present DIF, indicating that an item is invariant across different groups (gender, race, and cultural background). Using Mantel-Haenszel's procedure to analyze the items one at a time, researchers can obtain evidence of the interaction between items and person's characteristics, i.e., violation of invariance property [31].

Study Goals: The Rasch model is generally constructed to answer fundamental questions such as, "How do the assessment items represent an underlying construct?" and "To what extent is the construct validity of the assessment instrument established?". A construct is an ability possessed or endorsement supported by people. It is defined as an instrument being able to measure what it intends to measure [32, 33] Two important criteria of measurement science that need to be simultaneously considered are reliability and validity [34]. The validity focuses on the accuracy of the measurement while the reliability emphasizes the consistency or precision of the measurement, especially with repeated measures [33, 35]. The goals of this study are threefold: (1) to survey medical students throughout the curriculum for their perception regarding critical thinking attainment; (2) to evaluate the construct validity of the basic science curriculum assessment instrument, and (3) to determine areas of curricular modification and potential learning opportunities to maximize critical thinking abilities of medical students at the College.

Study Method: The development of the basic science curriculum assessment items included a review of critical thinking literature and Association of American Medical College (AAMC) graduating student surveys. The content validity of the assessment instrument was validated by an academic dean of the medical school who has tremendous experience with the curriculum. The formulated assessment instrument was then evaluated for its reliability, validity, and applicability using the Rasch model. The fit statistics (infit and outfit mean square residuals), separation index, unidimensionality, and invariance measurement were analyzed as a part of the assessment process. In the study, first- and second-year medical students (N1=109 and N2=98) as well as graduating medical students (N3=83) were asked to complete the paper-and-pencil based questionnaire. Approximately 45% (130) of 290 medical students in academic year 2016/17 completed a 22-item basic science curriculum assessment instrument (See Table 1). The response category option was the Likert measurement scale using a five-point rating with 0 being "Not Applicable", 1 being "Strongly Disagree", 2 being "Disagree", 3 being "Agree", and 4 being "Strongly Agree".

Table 1. Basic Science Curriculum Assessment Items

Item Name	Item Description
B1RESPON	The basic science curriculum remains responsive to feedback from students.
B2INNOVA	The basic science curriculum is open to innovation.
B3INTEGR	The basic science curriculum is well-coordinated and integrated.
B4MISSIO	The delivery of the basic science curriculum is coherent and compatible with the College's mission.
B5SMALL_	The basic science curriculum promotes small-group (8-12 students) teaching.
B6SCIENT	The basic science curriculum offers me adequate scientific knowledge.
B7SUB_BR	Courses in the basic science curriculum prepared me adequately for the Subject Board examinations.
B8USMLE1	Courses in the basic science curriculum prepared me adequately for the Step 1.

B9CLINIC	The basic science curriculum had clinical relevance contributing to my competency attainment in medicine.
B10INSIG	The basic science curriculum helped me gain insight into the disease processes.
B11SES_T	The basic science curriculum helped me evaluate social and economic trends and their impact on healthcare.
B12IMPRO	The basic science curriculum helped me participate in improving the healthcare of individuals, families, and groups in the community.
B13PATHO	Courses in the basic science curriculum enhanced my ability to recognize important pathology and physiology concepts within a patient case.
B14CRITI	Courses in the basic science curriculum enhanced my ability to critique research papers from peer review journals.
B15STATI	Research requirements in the basic science curriculum helped me apply appropriate statistical methods to scientific research.
B16CTS_S	Basic science small group teaching helped me improve my critical thinking skills to solve healthcare problems.
B17INTG_	Basic science small group teaching helped me integrate concepts taught in basic science lectures.
B18TABLE	Basic science examination questions included data tables and graphs similar to those on NBME Subject Board examinations.
B19MULTI	Basic science examinations included questions requiring a multi-step thinking approach similar to those on NBME Subject Board examinations.
B20WRITE	The basic science curriculum enhanced my reasoning ability to write scientific research papers.
B21VERB_	The basic science curriculum enhanced my reasoning ability to verbally present my research results.
B22OVERA	Overall, I am satisfied with the basic science curriculum at the College.

2. RASCH ANALYSIS

Person-Item Map: As shown in Figure 1, the five most difficult items above the character "S" (plus one standard deviation) inclusive are: B1RESPON, B2INNOVA, B20WRITE, B3INTEGR, and B8USMLE1: This indicates that the College's challenging areas in medical curriculum from student perspectives are: (1) "Basic science curriculum remains responsive to feedback from students"; (2) "Basic science curriculum is open to innovation"; (3) "Basic science curriculum enhanced my reasoning ability to write scientific research papers"; (4) "basic science curriculum is well coordinated and integrated"; and (5) "Courses in the basic science curriculum prepared me adequately for the Step 1".

On the contrary, the five least difficult items below the character "S" (minus one standard deviation) inclusive are B10INSIG, B5SMALL, B13PATHO, B11SES_T, and B9CLINIC. This shows that the College's strengths in medical curriculum from student viewpoints are: (1) Basic science curriculum helped me gain insight into the disease processes; (2) Basic science curriculum promotes small-group (8-12 students) teaching; (3) "Courses in the basic science curriculum enhanced my ability to recognize important pathology and physiology concepts within a patient case"; (4) "Basic science curriculum helped me evaluate social and economic trends and their impact on healthcare"; and (5) "Basic science curriculum had clinical relevance contributing to my competency attainment in medicine".

Figure 1. Person-Item Map

```
MEASURE    PERSON - MAP - ITEM
             <more>|<rare>
   4               +
                   |
              .    |
                   |
   3         .# T+
              .    |
              #    |
             .#    |
   2        .## S+
          #####    |
        .######    |T
          #####    |
   1   .#######    +  B1RESPON  B2INNOVA
           #### M|S B20WRITE  B3INTEGR  B8USMLE1
          .####    |  B18TABLE  B22OVERA
          .####    |  B14CRITI  B16CTS_S  B19MULTI  B21VERB   B4MISSIO
   0        ###   +M B7SUB_BR
        .######    |  B12IMPRO  B15STATI  B17INTG
             ## S|   B6SCIENT
             ##   |S B11SES_T  B9CLINIC
  -1          .    +  B13PATHO
              #    |
              . T|T B10INSIG  B5SMALL
                   |
  -2               +
                   |
              .    |
                   |
  -3               +
              .    |
                   |
                   |
  -4               +
             <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

Item Fit Order: All 22 items are appropriate (no misleading or reversed items) since all correlation coefficients are positive and greater than the 0.3 criterion [36]. The item fit order shows the 22 item infit and outfit statistics. Almost all of the items have infit mean square values within the range of 0.6 to 1.4, suggesting that they fit the Rasch model well with the exception of three somewhat misfitting items: B5SMALL B12IMPRO, and B22OVERA have infit and outfit mean square values of 1.68 and 1.63 for B5SMALL; 1.48 and 1.46 for B12IMPRO; and .49 and .50 for B22OVERA, respectively.

Reliability and Separation Indices: In the summary statistics of the WINSTEPS 3.91 software, Cronbach α shows the internal consistency reliability. The Cronbach-α value of 0.97 is surprisingly high which is well above the acceptable level 0.6**.** The reliability and separation of both item difficulty and person's ability have different applications and implications. Higher values of item and person reliabilities show that items produce consistent results for student respondents and that persons endorse steady marks for item group. High values of item and separation indices indicate that items have a large difficulty range and that persons have a large sample of respondents.

In this study, 122 persons have a separation value of 2.89 (not less than 2) and reliability of 0.89 (not less than 0.8) while the 22 items have separation value of 4.97 (not less than 3) and reliability of 0.96 (not less than 0.9). The indices of person and item separation and reliability, along with individual criterion in the prentices, support evidence for precise measurement of the basic science curriculum assessment. In addition, all person and item have infit mean square values within the range from 0.6 to 1.4, maintaining that data fit the Rasch model well.

Unidimensionality: It is assessed based on residual-based principal components analysis (PCA) using WINSTEPS. It is unlikely that the unexplained variance in the first contrast will be greater than 2.0 based on the Rasch model simulations [37]. However, the strength of first contrast for unexplained variance is 3.08 (items) measuring by eigenvalue, which is greater than the criterion of 2.0 items. Therefore, the violation of unidimensionality assumption is confirmed.

Differential Item Functioning (DIF): Mantel-Haenszel chi-square test is designed to determine if the DIF or systematic item bias effect exists for the student respondent groups (first-year, second-year, and graduating students) among these items. The probability values of Mantel-Haenszel chi-square test statistics were less than the .05 significance level with the exception of seven items (B5SMALL, B7SUB_BR, B12IMPRO, B16CTS_S, B17INTG, B18TABLE, and B19MULTI), showing this curriculum assessment instrument was systematically biased toward first- and second-year medical students as well as graduating medical students. Therefore, data analysis provided evidence of the violation of invariance property in the Rasch model, suggesting that the basic science curriculum assessment should be separately analyzed for first-year, second-year, and graduating medical students, respectively.

3. DISCUSSION

The study accomplished its objectives of assessing the construct validity of the basic science curriculum assessment instrument, and identifying the strengths and challenges of the basic science curriculum. The research findings showed that the person and item separation indices were large enough to adequately represent the spread or separation of persons or items on the rating response scale. Thus, the Rasch model has demonstrated internal consistency (high reliability), proving it to be suitable for assessing the medical curriculum.

The strengths of the basic science curriculum from the students' viewpoint were discovered, including the following components: "Basic science curriculum helped me gain insight into the disease processes"; "Courses in the basic science curriculum enhanced my ability to recognize important pathology and physiology concepts within a patient case"; and "Basic science curriculum helped me evaluate social and economic trends and their impact on healthcare". These positive findings are promising, as the perceived strengths of the curriculum are related to important skills (recognizing concepts within a case), and the application of concepts (understanding disease processes and socioeconomic determinants of healthcare) required of a medical professional. Adult learning theory which is currently being examined to enhance the learning and instruction of critical thinking in the medical curriculum, exhibits teaching information in a similar format and context to demonstrate its usefulness in practice to greatly enhance retention and application. Instruction in this manner also increases student motivation for learning [38, 39, and 40].

Some of the perceived challenging areas in the basic science curriculum according to the student assessment included: "Basic science curriculum is open to innovation"; "Basic science curriculum is well coordinated and integrated"; and "Courses in the basic science curriculum prepared me adequately for the Step 1". In order to improve some of these challenging areas, the medical curriculum needs to be adjusted in terms of integration of critical thinking skills throughout all four years of learning, and inclusion of skills for success on the USMLE Step examinations.

4. CONCLUSION AND IMPLICATION

Overall, the basic science curriculum assessment instrument provides the diagnostic feedback to not only establish the reliability and validity of the instrument, but also improve the effectiveness of the basic science portion of the curriculum, particularly as the entire curriculum is being reviewed for revision. In this study, only three items marginally exceeded the values of Infit and Outfit statistics, including: "Basic science curriculum promotes small-group (8-12 students) teaching."; "Basic science curriculum helped me participate in improving the healthcare of individuals, families, and groups in the community"; and "Overall, I am satisfied with the basic science curriculum at the College". In the foreseeable future, it would be interesting to see the responsiveness and sensitivity of the instrument change with the deletion of items. Also, performing an anchoring procedure by temporarily removing a few extreme persons after the initial analysis would further improve the construct validity of this assessment instrument.

However, when the principal component analysis (PCA) was used to evaluate the unidimentionality of the assessment instrument, the eigenvalue of 3.08 was discovered in the first contrast, indicating that approximately three items were measuring an alternative construct. To make the judgment about the unidimensionality of the assessment instrument, researchers examined the content of these items to see if they were related to different content as the sign of the multidimensionality. As a result, there was meaningful difference in the item content to support the multidimensionality. Therefore, adding valid items and removing unrelated items are needed to establish the construct validity of the assessment instrument for future application in the College.

5. REFERENCES

[1] R.M. Epstein, Assessment in Medical Education, **N Engl J Med**, 356:387-396, 2007, DOI: 10.1056/NEJMra054784.

[2] R**.** Speyer, W. Pilz, J. Van Der Krus, et al, Realibility and Validity of Student Peer Assessment in Medical Education: A Systematic Review, **Medical Teacher** 33: e 572-e585, 2011.

[3] C.K. Chen, Curriculum Assessment Using Artificial Neural Network and Support Vector Machine Modeling Approaches: A Cadse Study, **IR Applications**, 2010.

[4] United States Medical Licensing Examination, USMLE Content Outline, 2016, http://www.usmle.org/pdfs/usmlecontentoutline.pdf, Retrieved February 10, 2017.

[5] P.A. Facione, Critical Thinking : A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction Executive Summary, "The Delphi Report", **Calif Acad Press**, 1990, 423(c):1-19, doi:10.1016/j.tsc..07.002, 2009.

[6] G.M. Oderda, R.M. Zavod, J.T. Carter, et al, An Environmental Scan on the Status of Critical Thinking and Problem Solving Skills in Colleges/Schools of Pharmacy: Report of the 2009-2010 Academic Affairs Standing Committee, **Am J Pharm Educ**.,74(10):S6, doi:10.5688/aj7410S6, 2010.

[7] E. Krupat, J.M. Sprague, D. Wolpaw, P. Haidet, D. Hatem, B. O'Brien, Thinking Critically About Critical Thinking: Ability, Disposition or Both?, **Med Educ**., 45(6):625-635, doi:10.1111/j.1365-2923.2010. 03910.x, 2011.

[8] L. Bao, T. Cai, K. Koenig, et al, PHYSICS: Learning and Scientific Reasoning, **Science,** 323(5914):586-587, doi:10.1126/science.1167740, 2009.

[9] J. Stiller, S. Hartmann, S. Mathesius, et al, Assessing Scientific Reasoning: A Comprehensive Evaluation of Item Features That Affect Item Difficulty, **Assess Eval HighEduc**., 41(5):721-732, doi:10.1080/02602938.2016. 1164830,2016.

[10] O.L. Liu, L. Frankel, K.C. Roohr, Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment, **Educ Test Serv,** Princeton, NJ;(June):1-23, doi:10.1002/ets2.12009, 2014.

[11] W.C. Cox, J.E. McLaughlin, Association of Health Sciences Reasoning Test Scores with Academic and Experiential Performance, **Am J Pharm Educ**., 78(4), doi: 10.5688/ajpe78473, 2014.

[12] C.D. Kreiter, G. Bergus, The Validity of Performance-Based Measures of Clinical Reasoning and Alternative Approaches, **Med Educ**., 43(4):320-325, doi:10.1111/j.1365-2923.2008.03281.x, 2009.

[13] A. Power, J.F. Lemay, S. Cooke, Justify Your Answer: The Role of Written Think Aloud in Script Concordance Testing, **Teach Learn Med.**, 29(1):1-9. doi:10.1080/10401334.2016.1217778, 2016.

[14] R.G. Williams, D.L. Klamen, C.B. White, et al, Tracking Development of Clinical Reasoning Ability Across Five Medical Schools Using a Progress Test, **Acad Med.**, 86(9):1148-1154, doi:10.1097/ACM.0b013e31822631b3, 2011.

[15] K. Huhn, L. Black, G.M. Jensen, J.E. Deutsch, Construct Validity of the Health Science Reasoning Test, **J Allied Health**, 40(4):181-186, 2011.

[16] Cultivating Students' Critical Thinking at Meharry Medical College, 2017, Quality Enhance Plan Internal document.

[17] K. Royal, Making Meaningful Measurement in Survey Research: A Demonstration of the Utility of the Rasch Model, **IR Applications,** 28:1-16, 2010.

[18] D. Andrich, A Rating Formulation for Ordered Response Categories, **Psychometrika,** 43, 561-73, 1978.

[19] B.D. Wright, M.M.C. Wok, An Overview of the Family of Rasch Measurement Models, In E. V. Smith, Jr. & R.M. Smith, **Introduction to Rasch Measurement: Theory, Models, and Application**, Pages 1-24, 2004.

[20] T.G. Bond, C.M. Fox, **Applying The Rasch Model: Fundamental Measurement In The Human Sciences,** (2nd ed.), Mahwah, N.J.: Lawrence Erlbaum Associates Publishers, 2007.

[21] P.T. Oon, R. Subramaniam, Factors Influencing Singapore Students' Choice of Physics as a Tertiary Field of Study: A Rasch Analysis, **International Journal of Science Education**, Volume 35, 2013 - Issue 1, Pages 86-118.

[22] P.D. Hart, M. Kang, N.L. Weatherby, Y.S. Lee, T.M. Brinthaupt, Evaluation of the Short-Form Health Survey (SF-36) Using the Rasch Model, **American Journal of Public Health Research,** 2015, Vol. 3, No. 4, 136-147 Available online at http://pubs.sciepub.com/ajphr/3/4/3 © Science and Education Publishing DOI:10.12691/ajphr-3-4-3.

[23] H.H. Liu, Y.S. Lee, Measuring Self-regulation in Second Language Learning: A Rasch Analysis on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

[24] J.M. Linacre, A User's Guide to WINSTEPS, Chicago, IL., Winsteps.com. Retrieved February 10, 2017, Available from http://www.winsteps.com/.

[25] B.D. Wright, J.M. Linacre, J.E. Gustafson, P. Martin-Lof, Reasonable Mean-Square Fit Values, **Rasch Measurement Transactions**, 8(3), 370, 1994.

[26] S. Gracia, Analyzing CSR Implementation with the Rasch Model, **Faculty Publications,** Paper 271, 2005.

[27] M. Wu, R. Adams, **Applying the Rasch Model to Psycho-social Measurement: A Practical Approach, Educational Measurement Solutions**, Melbourne, 2007.

[28] J.M. Linacre, Winsteps® (Version 3.72.0) [ComputerSoftware, 2011], Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2011, Available from http://www.winsteps.com/.

[29] A. Tennant, S.P. McKenna, P. Hagell, Application of Rasch Analysis in the Development and Application of Quality of Life Instruments, **Value in Health**, 7 (2004), S22–S26.

[30] A. Tennant, M. Pent, L. Tesio, G. Grimby, J.L. Thonnard, A. Slade, et al., Assessing and Adjusting for Cross-Cultural Validity of Impairment and Activity Limitation Scales Through Differential Item Functioning Within the Framework of the Rasch Model: the PRO-ESOR Project, **Medical Care**, 42 (1 Suppl.) 2004, pp. I37–I48.

[31] Mantel-Haenszel Procedure (1959, www.rasch.org/ memo39.pdf, RMT 1989 3:2 51-53).

[32] J.D. Brown, What is Construct Validity? **JALT Testing & Evaluation SIG Newsletter,** 2000; 4(2): 8-12.

[33] R.H. Friis, T.A. Sellers. Chapter 11: Screening for Disease in the Community, **Epidemiology for Public Health Practice**, United States of America: Jones and Bartlett Publishers, 2009: 409-36.

[34] M. Tavakol, R. Dennick, Making Sense of Cronbach's Alpha, **Int J Med Edu**, 2011, 2: 53-5.

[35] A. Aron, E.N. Aron, E.J. Coups, **Statistics for Psychology** (5th ed.), USA:Pearson Prentice Hall, 2009.

[36] J.M. Linacre, Winsteps® Ministep, **Rasch-Model Computer Programs, Program Manual** 3.91.0, 2015, Available from: http//www. winsteps.com.

[37] J.M. Linacre, Winsteps®, **Rasch Measurement Computer Program User's Guide**, 2016, Beaverton, Oregon, Available from http://www.winsteps.com, Retrieved February 10, 2017.

[38] D.G. Brauer, K.J. Ferguson, The Integrated Curriculum in Medical Education: AMEE Guide No. 96, **Med Teach**, 2015;37(4):312-322, doi:10.3109/0142159X. 2014.970998.

[39] D.C.M. Taylor, H. Hamdy, Adult Learning Theories: Implications for Learning and Teaching in Medical Education: AMEE Guide No. 83, **Med Teach**. 2013;35(11):e1561-72, doi:10.3109/0142159X, 2013. 828153.