

women younger than 30 years of age are moderate. The harms include more frequent testing and invasive diagnostic procedures such as colposcopy and cervical biopsy. Abnormal screening test results are also associated with psychological harms, anxiety and distress.

"It has been estimated that nearly one-third of health care delivered in the United States is unnecessary," said Stream. "Tests and procedures that lack evidence of their effectiveness put our patients at risk and drive up the cost of care."

To date, more than 130 questionable tests and procedures have been released as part of the Choosing Wisely campaign. The organizations that joined the AAFP in releasing this latest round of recommendations are the American Academy of Hospice and Palliative Medicine; American Academy of Neurology; American Academy of Ophthalmology; American Academy of Otolaryngology-Head and Neck Surgery; American Academy of Pediatrics; American College of Obstetricians and Gynecologists; American College of Rheumatology; American Geriatrics Society; American Society for Clinical Pathology; American Society of Echocardiography; American Urological Association; Society for Vascular Medicine; Society of Cardiovascular Computed Tomography; Society of Hospital Medicine; Society of Nuclear Medicine and Molecular Imaging; and the Society of Thoracic Surgeons.

A third wave of lists will be unveiled later in 2013, including another 5 recommendations by the AAFP. For that round, the Academy is expected to be joined by the AMDA—Dedicated to Long Term Care Medicine; American Academy of Dermatology; American Academy of Orthopaedic Surgeons; American College of Chest Physicians; American College of Rheumatology; American College of Surgeons; American Headache Society; American Society for Radiation Oncology; American Society of Clinical Oncology; American Society of Hematology; American Thoracic Society; Heart Rhythm Society; North American Spine Society; and the Society of General Internal Medicine.

Matt Brown
AAFP News Now



From the American
Board of Family Medicine

Ann Fam Med 2013;11:286-288. doi:10.1370/afm.1536.

DIMENSIONALITY OF THE MAINTENANCE OF CERTIFICATION FOR FAMILY PHYSICIANS EXAMINATION: EVIDENCE OF CONSTRUCT VALIDITY

The American Board of Family Medicine (ABFM) Maintenance of Certification for Family Physicians (MC-FP) examination is designed to measure a single construct: clinical decision-making abilities within the scope of practice of family medicine. Implied in the construct of clinical decision-making abilities is the ability to recall relevant elements from a large fund of pertinent medical knowledge. While clinical decision-making abilities could be perceived as comprising several separate constructs (eg, based upon clinical categories, organ systems, etc), that approach would require the development of multiple assessment scales with a passing criteria specific to each. Instead, the overarching construct of clinical decision-making ability, which encompasses those more specific areas, has been selected by the ABFM because it more closely mirrors the pass-fail decision process used to discern which candidates receive certification. In any instance, the construct that the ABFM attempts to measure needs to be sufficiently unidimensional in order to produce precise, error-free estimates of a candidate's performance. This brief article will discuss the dimensionality of the MC-FP examination and its implications for construct validity, namely the validation that the examination accurately measures the ability of family physicians to make appropriate clinical decisions.

Dimensionality

Why is dimensionality important? Simply put, it is desirable to measure only 1 thing at a time. Just as physical measurement attempts to measure 1 thing at a time (eg, a patient's blood pressure reading should not be biased by his/her height, weight, or sex), psychometricians, the measurement experts that help design our examinations, also aspire to measure only 1 latent trait at a time. It is only when dimensions are clearly isolated that one can understand the meaning of the measure and make a valid inference about an examination score.

Dimensionality of the MC-FP Examination

As we have mentioned previously, the psychometric model that the ABFM employs to score its examinations is the Rasch model, a 1-parameter Item Response Theory (IRT) measurement model. This model converts raw scores to linear measures and controls for the difficulty of the test version a candidate received.¹ In addition to using typical fit indicators, the most effective way to detect multidimensionality in a Rasch measurement-based data analysis is to use a Principal Components Analysis (PCA) of standardized residual correlations.² In short, the Rasch model uses ordinal data to construct a one-dimensional measurement system. Of course, real data are never perfectly unidimensional, so the presence of more than 1 latent dimension in the data always exists to some extent. When the data perfectly fit the Rasch model (this includes all items and persons examined) all systematic variation is explained by a single dimension. Data that do not perfectly accord with the model leave behind residuals that have a random normal structure and predictable variance.²

To evaluate the dimensionality of the MC-FP examination, we perform the aforementioned industry standard tests of fit and PCA of standardized residual correlations. An investigation of data-to-model fit, both overall and by individual item analysis, can help us discern if multiple dimensions are present and exactly where these dimensions might be in the dataset. To demonstrate this, let us share an analysis we performed using the core portion of the 2010 examination. The dataset included 3,697 examinees and the 423 test items that appeared across the multiple forms of the core portion of the MC-FP examination. Fit statistics indicated perfect overall data-to-model fit, with infit and outfit mean square statistics of 1.0 for both persons and items. Values of 1.0 are ideal for these analyses,³ and the acceptable range is between 0.80 and 1.20.⁴ Individual item fit statistics were then evaluated. Only 8 of 423 items deviated from the ideal range. The most over-fitting item had a mean square value of 1.27, and the most under-fitting item had a mean square value of 0.77. Meaning, less than 2% of the items appearing on the MC-FP examination had fit statistics that fell outside the ideal range for dichotomous data. These statistics indicate excellent item fit with minimal off-variable noise.

Next, the slight noise that was detected in the measures was evaluated by way of a PCA of standardized residual correlations. The candidates who complete the MC-FP examination each year are quite homogeneous; they are highly educated physicians with expertise in family medicine. Therefore, a great deal of variability across person measures (mean score, 469; SD, 98) and item measures (mean score, 297; SD, 168) does not

exist, considering the reported range of scores is from 200 to 800. Naturally, this lack of variation leads to an inability to explain a great deal of the variance.⁵ Data from this MC-FP examination explained just 11.2% of the variance. The vast majority of variance (7.5%) explained came from the test items. The strongest secondary dimension detected explained 1.2% of the variance. The ratio between the overall primary dimension and the secondary dimension was 11.2:1.2; the ratio between the primary item dimension and the strongest secondary dimension was 7.5:1.2. These ratios are universally accepted in the measurement literature as being sufficiently unidimensional.^{6,7}

The most polarizing items that appeared on the examination from a dimensionality perspective were identified by the PCA analysis and reviewed by content experts. The nature of these items pertained to issues of prevention at one extreme and issues of treatment at the other. The items underwent a psychometric evaluation, and all psychometric indicators confirmed the items functioned properly and were indeed good, quality items. Because family physicians are expected to be knowledgeable of both the prevention and the treatment of illnesses, the substantive nature of the detected secondary dimension appeared, therefore, to be rather inconsequential.

The MC-FP examination is intended to measure the single construct of clinical decision-making ability within the practice of family medicine. Results of the dimensionality analysis described above indicated the MC-FP examination is highly unidimensional from a psychometric perspective. That is, the data accorded well with the model's expectations and the internal structure of the data was correlated in such a way that the same construct was being consistently measured throughout the examination. Expert content review of the substantive content of polarized dimensions provided additional assurance of the unidimensional nature of the examination.

What do these results mean with regard to examination score validity? Renowned measurement scholar Samuel Messick conceptualized construct validity as a uniform concept that required multiple pieces of evidence.⁸ He identified 6 aspects of construct validity: content, substantive, structural, generalizable, external, and consequential. When evaluating the results of the analysis of our examination from Messick's framework, psychometric evidence is available that speaks to the content, substantive, and in some limited way, structural aspects of construct validity. We have previously provided some evidence that speaks to the generalizable aspect of validity as well.⁹ Collectively, these results should be reassuring for candidates, as they provide additional evidence of the psychometrically sound

nature of the MC-FP examination. Of course, test takers also should be assured that the MC-FP examination yields valid inferences about their scores as well.

Kenneth D. Royal, PhD; James C. Puffer, MD

References

1. Royal KD, Puffer JC. Understanding the "sum of subtest to overall score discrepancy" on the maintenance of certification for family physicians examination. *J Am Board Fam Med*. 2012;25(2):260-261.
2. Linacre JM. Detecting multidimensionality: which residual data-type works best? *J Outcome Meas*. 1998;2(3):266-283.
3. Linacre JM, Wright BD. Dichotomous mean-square infit and outfit Chi-square fit statistics. *Rasch Meas Trans*. 1994;8(2):360.
4. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans*. 1994;8(2):370.
5. Linacre JM. Variance in data explained by Rasch measures. *Rasch Meas Trans*. 2008;22(3):1164.
6. Linacre JM. Data variance: Explained, modeled and empirical. *Rasch Meas Trans*. 2003;17(3):942-943.
7. Linacre JM. Dimensionality: Contrasts and variances. *Winsteps Manual*. 2011. <http://www.winsteps.com/winman/index.htm?principalcomponents.htm>. Accessed Mar 2, 2012.
8. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol*. 1995;50(9):741-749.
9. Royal KD, Puffer JC. The reliability of American Board of Family Medicine examinations: implications for test takers. *J Am Board Fam Med*. 2012;25(1):131-133.

available at <http://www.fammedrcr.org>. The resource currently includes a core topic list for postgraduate year one (PGY-1) with recommended readings. Recommended readings include evidence-based articles and clinical practice guidelines. "Programs can use what's on the site now to identify what should be taught during the first year of residency and quickly access the latest evidenced-based review article," said Dr Tuggy. Access to this information is available at no charge to STFM and AFMRD members.

The Family Medicine Residency Curriculum Resource is a multiyear collaborative project of The Association of Family Medicine Residency Directors (AFMRD) and Society of Teachers to build an online resource of peer-reviewed, competency-based curriculum content organized by post-graduate year (PGY). When the site is complete, it will house peer-reviewed, competency-based curricular resources covering the core content of family medicine education. Subscribers will be able to download case-based presentations, quizzes, and facilitators' guides for each core topic.

The site is being developed in phases:

- October 2012: Call for Proposals for PGY-1 Curriculum
- April 2013: Launch PGY-1 topic list and recommended readings
- December 2013: Launch PGY-2 and PGY-3 topic list and recommended readings
- April 2015: Launch subscriber-only curricular resource

The editorial team includes Dr Tuggy as senior editor and Julia Fashner; Karen Hall, MD; Melissa Stiles, MD; Tochi Iroku-Malize, MD, MPH; Doug Maurer, DO, MPH; and Jenifer Van Deusen, MEd as associate editors.

Dr Tuggy pointed out that the resource can only be successful through collaboration. "STFM and AFMRD have funded the architecture and development of the site. We now need each residency program to step up and submit curriculum for at least one topic."

Visit <http://fammedrcr.org/rcrnew/call-for-proposals> for a proposal application and a list of available topics.

Mary Theobald, MBA

*Vice President of Communications and Programs
Society of Teachers of Family Medicine*



Ann Fam Med 2013;11:288. doi:10.1370/afm.1546.

FAMILY MEDICINE RESIDENCY CURRICULUM RESOURCE IS NOW LIVE

Mike Tuggy, MD, has been envisioning this for years: a shared, online repository of high-quality residency curriculum created by residency programs. "We've got more than 450 family medicine residency programs each creating curricula for more than 400 core clinical topics. If we collaborate, we can save hundreds of hours a year and also improve the quality of our teaching."

Dr Tuggy's vision is taking shape in the just-released *Family Medicine Residency Curriculum Resource*