Dynamic Measurement in Health Professions Education: Rationale, Application, and Possibilities

Denis Dumas, PhD, Daniel McNeish, PhD, Deanna Schreiber-Gregory, MS,
Steven J. Durning, MD, PhD, and Dario M. Torre, MD, PhD, MPH

Abstract

Dynamic measurement modeling (DMM) is a psychometric paradigm that uses longitudinal data to estimate individual students' growth in measured skills over the course of an educational program (i.e., growth scores). DMM represents a more formal way of assessing learning progress across the health professions education continuum. In this article, the authors provide justification for this approach in health professions education and demonstrate its proof-of-concept use with three time points of United States Medical Licensing Examination Step exams to generate growth scores for 454 current and recent medical learners. The authors demonstrate that learners vary substantially on their growth scores, and those growth scores exhibit psychometric reliability. In addition, growth scores significantly and positively correlated with indicators of medical learner readiness (e.g., undergraduate grade point average and Medical College Admission Test scores). These growth scores were also capable of significantly and positively correlating with future ratings of clinical competencies during internship as assessed through a survey sent to their program directors at the end of the first postgraduate year (e.g., patient care, interpersonal skills). These preliminary findings of reliability and validity for DMM growth scores provide initial evidence for further investigation into the suitability of a dynamic measurement paradigm in health professions education.

Today, most health professions educators understand that trainees differ in their learning trajectories. For example, it is possible for some learners who perform comparatively worse than their peers on an assessment in the beginning of their training to catch up, or vice versa, learners who perform comparatively better than their peers initially may subsequently slow their learning and therefore fall behind.[1] This widely observed phenomenon demonstrates that the comparative rank-order of learners on any given measured skill is not fixed throughout their training or clinical experience: The growth trajectory of any measured ability can have a different slope (and possibly a different shape altogether) across different learners. Therefore, students' scores on a single-time-point test are merely a quantitative snapshot of their current knowledge and ability at the time of testing, and it cannot be known from a single static score whether a learner is improving or declining relative to their peers.

By way of analogy, health professionals understand the importance of growth curves in pediatrics because a child's growth and development over time are much more informative than their data from any individual checkup. In the same way, we contend that learning trajectories modeled over time are much more informative about trainees' academic development than are any of their individual test scores. Unfortunately, static test scores give essentially no information about a learner's growth over time.[2] We believe that such a situation is problematic for the health professions field because, in an era of rapidly changing technological and social contexts—as well as the perennial need for health professionals to effectively and rapidly discern causes of a patient's presentation in the clinical context—a physician's (or medical learner's) most highly relevant mental attribute may be their ability to learn and adapt to new information.[3,4]

To help rectify this issue, some institutions have developed progress-testing programs, which often feature all learners being administered the same test over time, with the expectation that entering trainees will initially perform very poorly, with their performance improving and approaching mastery of the material as they move through their schooling.[5–8] In such a progress-testing program, if all learners' scores were plotted over time for every testing occasion, the various growth trajectories of the learners would be discernable, and some rough determination of an individual learner's skill improvement would be estimable. Unfortunately, because current progress-testing programs are not based on a formalized psychometric framework for estimating student growth trajectories, at this time the calculation of reliability and validity indices associated with progress-testing scores is not possible. Therefore, the amount of measurement error in progress-testing growth estimates is currently unknown, making the use of these progress trajectories impossible for any context in which the reliability of findings is of critical importance, whether those be research or practical assessment scenarios. Further, the statistical methods currently used in progress testing (i.e., general linear modeling[9]) make the same assumption of rank-order preservation among learners as traditional static testing.

For this reason, progress testing has nearly exclusively been applied in

Please see the end of this article for information about the authors.

Correspondence should be addressed to Denis Dumas, Department of Research Methods and Information Science, University of Denver, Denver, CO 80208; email: Denis.Dumas@du.edu.

Acad Med. 2019;94:1323–1328.
First published online April 2, 2019
doi: 10.1097/ACM.0000000000002729

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A663 and http://links.lww.com/ACADMED/A664.

medical education only as a low-stakes assessment,[6,10] and findings from progress tests have not as of yet been applied to higher-stakes decisions about learners. Such a situation is unfortunate because the learning trajectory of a trainee, like a growth curve for a child, can provide important information for decisions ranging from residency placement to board licensure. Here, we propose a recently developed psychometric framework—dynamic measurement modeling (DMM)—as a possible method for formalizing the measurement properties of trainee learning trajectories, as a way to support the development of longitudinal teaching and assessment paradigms within health professions education (HPE). In this article, we provide evidence of proof-of-concept in HPE, a demonstration that is meant to exhibit the possibilities and promise of DMM as a way to improve measurement practice within our field.

Dynamic Measurement Modeling

In contrast to the general progress-testing framework that exists within the medical education literature, a dynamic measurement paradigm for testing is a more specific endeavor oriented purposefully toward generating scores for every trainee that reliably and validly indicate that trainee's growth within the domain being assessed. In this way, DMM can be conceptualized as a psychometrically formalized longitudinal (e.g., progress) testing framework, in which the measurement error in growth trajectories is estimable. A dynamic testing paradigm has traditionally been applied in the identification of learning disabilities[11] and giftedness[12] in young children, but more recently, DMM has been applied to large-scale U.S. longitudinal data from kindergarten (K) through eighth grade to estimate students' growth in mathematics[13] and reading[14] over time. To date, applications of DMM are typically preliminary or research oriented in nature (rather than high-stakes). However, with the application of novel statistical techniques for estimating growth scores that have appeared since 2015,[2,13] combined with the already rich tradition of longitudinal testing in HPE,[5–8] we believe that HPE may be the ideal sphere of education to lead the field in application of a dynamic measurement paradigm.

In dynamic measurement, longitudinally scaled tests are administered to participants over time with instruction interspersed between the testing occasions. In this way, differential growth rates among learners can be quantitatively modeled, and scores that summarize learners' growth can be estimated.[2] A dynamic measurement model is a specialized class of mixed-effects model, and specific equations that were used here to formulate the model are presented and explained in Supplemental Digital Appendix 1, available at http://links.lww.com/ACADMED/A663. After fitting a dynamic measurement model, individual growth trajectories for each learner in the model can be used to create scores that summarize that learner's improvement on their skills over time. Although this approach is statistically complex, DMM is likely much more useful than more simplistic methods for inferring a learner's normative growth (e.g., by examining their percentile ranks over time) because DMM is capable of generating scores for every learner that represent the slope of their individual improvement over time, no matter how large the dataset, and those scores can be formally psychometrically examined for reliability. In our previous work,[13,14] such a quantity was also shown to be relatively unaffected by student race, gender, or socioeconomic status: a very important concern for the fairness and validity of educational measurement.

Empirical Proof-of-Concept Demonstration: United States Medical Licensing Examination Step Exams

To demonstrate the feasibility of adopting a DMM paradigm within the context of HPE, we used United States Medical Licensing Examination (USMLE) Step exam scores from 454 current and recent medical students at the Uniformed Services University of the Health Sciences in Bethesda, Maryland. This study was approved by the institutional review board at this institution. The study sample was 72.5% male (n = 329) and 27.5% female (n = 125). The average age at the time of medical school matriculation in this sample was 24.50 years (SD = 3.55). Of the 454 learners included in the sample, 395 (87.0%) reported their race/ethnicity, with these being demographically composed of 74.7% European American/white (n = 295), 20.2% Asian (n = 80), 2.2% African American/black (n = 9), 1.7% Hispanic/Latino (n = 7), and 1.0% Native American/other (n = 4).

Because every student in this sample completed all three USMLE Step exams, this analysis featured no missing data. It should be noted here that this DMM model was not, and is currently not recommended to be, used for high-stakes decisions about students. Instead, our goal in this analysis is to demonstrate that DMM may produce reliable and valid estimates of student learning and improvement, referred to here as *growth scores*, that may reasonably, after further careful consideration, be used for understanding student learning trajectories in contexts in which scores must be estimated with high precision, whether those contexts be focused on practical assessment or research.

Fitting the dynamic measurement model

To fit the model, we organized student scores on each of the three USMLE Step exams in a long-form dataset. DMM requires longitudinal data that are vertically scaled (i.e., growth occurs along a single vertical dimension across time). In the wider educational research literature, vertical scaling methods (almost always based on item-response theory) are sometimes applied to longitudinally administered tests to ensure that these scores can be placed on a single vertical scale. However, we are unaware of any sequence of measures that have been fully vertically scaled. Therefore, we took a pragmatic approach to modeling trainee growth and made the psychometric assumption that the material on each subsequent Step exam conceptually builds on the previous step (a conceptualization that is not uncommon in the literature[15]). Therefore, Step exam scores were additively combined such that Steps 1, 2, and 3 had a linear relationship across time along a single vertical scale. In longitudinal and growth modeling studies within the medical education literature, if vertically scaled measures become available, such an assumption may become obsolete. But given the currently available longitudinal scores, this assumption is necessary to model trainee growth in HPE. As another option, data from instances where trainees complete the same measure

serially over the course of a program, such as in a traditional progress-testing program, may also be appropriate for DMM, but the wide relevance of USMLE Step exam scores to HPE, in our view, made them an interesting proof-of-concept case for DMM in this context.

From this linear dynamic measurement model, we computed growth scores that represented individual students' learning slopes over time. Individual learning trajectories, estimated via this model, for all 454 students in the sample are plotted in Figure 1, with the mean trajectory of this sample plotted in bold. After computing the growth scores from the dynamic measurement model for each learner in our analytic sample, a histogram was plotted of the growth scores (Figure 2). A growth score of zero indicates average growth, and all nonzero growth scores (i.e., deviations from average) are on the scale of the USMLE Step exams. So, a learner with a growth score of 25 improved 25 Step exam points more than the average level of improvement (for all learners) between each of the three exams. In the same way, a learner with a growth score of −25 improved 25 points less than the average improvement between each of the three time points. In our view, growth scores on this scale may provide health professions educators with important insight into a particular learner's improvement during medical school, especially in cases where growth scores are below average and remediation might be warranted. Of course, it may be desired in future applications of DMM to track the growth of only a subset of learners (e.g., struggling learners) from a given HPE program. In those cases, growth scores would be interpreted as normatively pertaining to that sample specifically, and the mean growth score would likely not signify the same amount of growth as the mean growth score for a general sample. For this reason, dynamic measurement models with the widest sampling frame would capture the fullest distribution of growth scores and lead to the richest inferences about individual learners.

In this framework, a high score on the USMLE Step 1 exam would not necessarily produce a "ceiling effect" leading to a low growth score. In fact, just the opposite is more commonly observed in educational growth studies: Those

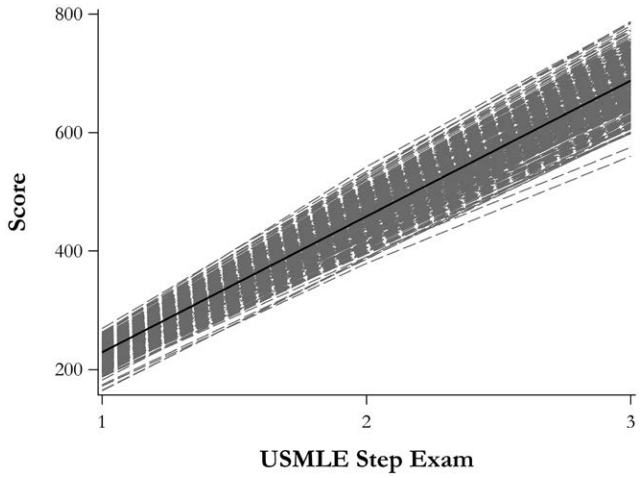

Figure 1 Linear dynamic measurement model growth trajectories for medical skills as assessed by the United States Medical Licensing Examination Step exams in the analytic sample of 454 medical students. All 454 individual growth trajectories are pictured. The mean trajectory is superimposed in bold.

learners who are ahead of their peers at the initial time point of measurement may tend to grow faster than those who are behind. This observation leads to an effect in which students who are behind initially fall further and further behind over time as a result of their slower growth. In this particular sample of medical trainees, such a "rich get richer" effect is also observed. Those trainees who scored highest on the first exam were most likely to receive a very high growth score, as indicated by the correlation between the initial Step exam score and the growth scores ($r = 0.95$). For this reason, the phenomenon of "regression toward the mean" over time for the highest scorers on the initial Step exam appears to be very unusual in this sample of trainees. It should be noted that, because DMM estimates learner-specific growth trajectories via a mixed-effects framework, DMM is useful whether a trainee approaches or departs from the mean over time, allowing for a flexible measurement framework that can account for a multiplicity of learning trajectories and growth scores.

It should also be noted here that, because the growth scores produced by a dynamic measurement model are inherently normative in nature, the placement of "cut scores" or standards that trainees must fall short of on the otherwise continuous growth distribution in

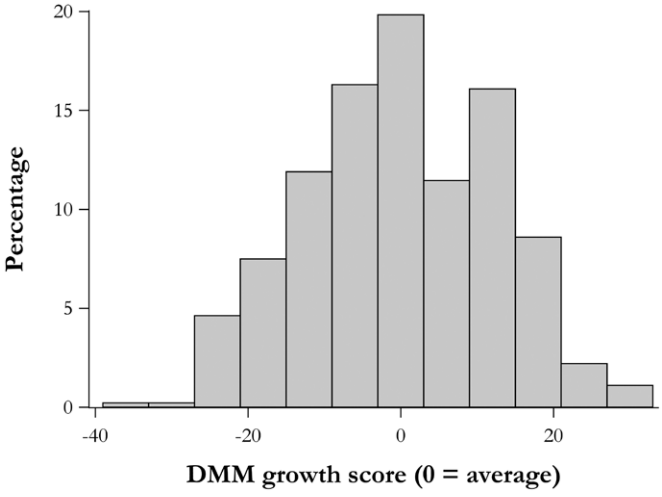

Figure 2 Histogram of dynamic measurement model estimated growth scores in the analytic sample of 454 medical learners. Scores are standardized such that zero is the sample mean growth score, and growth score units are scaled to be the same as USMLE Step exam scores. Abbreviation: USMLE indicates United States Medical Licensing Examination.

order to be targeted for remediation would need to be determined based on future research or expert opinion within individual training programs. We do believe that DMM gives individual programs flexibility in making these decisions that leaders deem appropriate for their learners. Further, drawing standards within a continuous score distribution is not unique to the DMM context but is an inherent issue associated with nearly all psychometric testing within and beyond the HPE context.

Reliability of growth scores

To determine whether these DMM growth scores may be suitable in the future for making justifiably precise decisions about students, we conducted a reliability estimation. The modern reliability index used here is estimated via Gaussian quadrature, and specific equations that were used for that estimation are shown in Supplemental Digital Appendix 1, available at http://links.lww.com/ACADMED/A663. Such a reliability statistic is different from classical "internal consistency" indices of reliability (e.g., Cronbach alpha) because it is intended to capture the proportion of true score variance in the growth scores, not any of the individual Step exam scores.

In this case, the strong reliability result of 0.92 indicates that the DMM growth scores estimated here include a very small proportion of error variance and, therefore, would be likely to display consistency were they to be reassessed. Put another way, the reliability statistic calculated here indicated that 92% of the variance in DMM growth scores estimated in this investigation was true variance indicating trainees' true improvement and learning, while 8% of the variance in growth scores was error variance created by noise in the growth trajectories.

Validity of growth scores

DMM growth scores for our analytic sample of students were first correlated with a number of indicators of learning readiness that trainees bring with them to medical school, including their undergraduate grade point average (GPA) and Medical College Admission Test (MCAT) scores. Because these indicators of learner readiness were required for matriculation to medical school, all 454 learners in the analytical sample also had these scores (no data are missing). As can be seen from Table 1, total undergraduate GPA was the strongest correlate of growth scores, closely followed by the biology scale of the MCAT, despite the near-zero correlation between GPA and MCAT biology. Such a pattern of correlations implies that although GPA and MCAT scores may not be closely linked for a given entering student, both these indicators may provide information about their growth within medical school. Of the readiness indicators we included in this analysis, only the verbal and writing sections of the MCAT failed to correlate with DMM growth scores, implying that those skills may be less associated with medical school learning than are biology and physics. It should be noted that none of the correlations observed between the indicators of learner readiness and DMM growth scores were above 0.22, indicating that such readiness indicators (e.g., MCAT scores) do not strongly correlate with growth across three time points of the USMLE Step exams. In the future, early indicators of growth other than GPA and MCAT scores, that may better correlate with growth scores, may be identified.

Next, DMM growth scores for our analytic sample were correlated with ratings of learners' clinical competency outcomes, which were generated by their internship supervisors (postgraduate year 1 surveys) using a Likert-style rating scale with five scale points, during their internship training. Each of the clinical competency scores was generated using a principal-components analysis of the individual rating scales; we have previously gathered reliability and validity evidence for this instrument.[16,17] In each case, the first component of the scale was used to calculate learner scores on the clinical competency being assessed. The first component of each rating scale accounted for between 76% (for the systems-based practice scale) and 88% (for the medical knowledge scale) of the variance on that scale. The overall clinical competency scale was also completed by interns' supervisors and represents a general dimension of clinical practice. The average response rate on the postgraduate year 1 clinical outcome survey was 47% across each of the scales. All of the items from this clinical outcome survey are included in Supplemental Digital Appendix 2, available at http://links.lww.com/ACADMED/A664.

As can be seen in Table 2, growth scores were most strongly correlated with the medical knowledge and patient care clinical competency ratings, which is unsurprising given that medical knowledge has been found to be closely associated with the Step exams. However, other dimensions of clinical competence less associated with the Step exams were also positively and significantly correlated with growth scores, including communication and interpersonal skills as well as professionalism ratings, implying that DMM growth scores in medical school may also be associated with such "softer" skills later in the medical career. It should be noted that the highest correlation observed between

Table 1

Correlations Among Growth Scores and Indicators of Medical Learner Readiness

Variable	A: Growth scores	B: Total GPA	C: Science GPA	D: MCAT verbal	E: MCAT physics	F: MCAT biology	G: MCAT writing	H: MCAT total
A	1.00							
B	0.22[a]	1.00						
C	0.18[a]	0.22[a]	1.00					
D	−0.01	−0.09	−0.10[b]	1.00				
E	0.15[a]	−0.10[b]	−0.04	−0.03	1.00			
F	0.20[a]	−0.02	0.02	−0.03	0.23[a]	1.00		
G	0.04	−0.02	−0.04	0.11[b]	0.01	−0.10[b]	1.00	
H	0.17[a]	−0.12[b]	−0.06	0.51[a]	0.69[a]	0.62[a]	0.02	1.00

Abbreviations: GPA indicates grade point average; MCAT, Medical College Admission Test.
[a]$P < .01$.
[b]$P < .05$.

Table 2

Correlations Among Growth Scores and Indicators of Medical Learner Clinical Outcomes During Internship

Variable	A: Growth scores	B: Overall clinical competence	C: Patient care	D: Communication and interpersonal skills	E: Medical knowledge	F: Professionalism	G: Systems-based practice	H: Military unique practice
A	1.00							
B	0.24[a]	1.00						
C	0.35[a]	0.74[a]	1.00					
D	0.22[a]	0.75[a]	0.91[a]	1.00				
E	0.40[a]	0.76[a]	0.86[a]	0.79[a]	1.00			
F	0.26[a]	0.67[a]	0.79[a]	0.84[a]	0.69[a]	1.00		
G	0.18[b]	0.66[a]	0.88[a]	0.91[a]	0.77[a]	0.83[a]	1.00	
H	0.13	0.69[a]	0.76[a]	0.78[a]	0.58[a]	0.68[a]	0.80[a]	1.00

[a]$P < .01$.
[b]$P < .05$.

growth scores and clinical outcome ratings in this study was 0.40, indicating a moderate relation (not a strong relation) between growth on the USMLE Step exams in medical school and future clinical outcomes.

Discussion

We have demonstrated that individual trainees vary in their growth and improvement during medical school, leading to differential growth rates among trainees on their USMLE Step exams. If static scores from a single exam were used to make decisions about learners, differences in growth trajectory would not be considered, therefore potentially weakening the validity of inferences made about learners. DMM offers a potentially useful alternative to traditional psychometric modeling in HPE, allowing for growth scores that summarize a trainee's improvement over time to be estimated. In the context of the USMLE Step exams, such growth scores demonstrate reliability and preliminary validity both in terms of their correlations with indicators of learning readiness (e.g., undergraduate GPA, MCAT scores) and critical measures of clinical competency during internship (e.g., patient care, professionalism). Such preliminary predictive validity findings stand in contrast to earlier work that shed doubt on the ability of USMLE Step exam scores to correlate with clinical competence outcomes,[18] and therefore point to the possible appropriateness of DMM for medical residency selection decisions.

In addition to potentially enhancing the validity inferences for all learners, growth scores may be important for other reasons. For example, identifying growth rates that have slowed or flattened over time could be used to support health professional training and remediation efforts. Going further, such an approach could be used to assist in the detection of critical mental health issues that affect academic growth (such as burnout, depression, and risk of suicide). Health professions learners are generally such high achievers that using a failing grade alone to detect learners who may be struggling with one or more of these mental health concerns does not provide the potential sensitivity of DMM: much like a child who falls off their height and/or weight curve.

Moreover, DMM may be flexibly applied throughout a training program such that, if information about a learner's growth is potentially more important early in a program rather than at the end to identify trainees in need of remediation, assessments that capture their growth over that time frame could be specifically modeled. Conversely, DMM could also be applied over a longer time frame, or with many more measurement time points, to describe a more general aspect of trainee growth. In particular, it is possible that nonlinear growth trajectories would emerge if trainees' skills were measured over a longer period of time—something that has been documented in K–12 educational research[2] and that may be interesting for future investigations into learning in the health professions. In this way, DMM is analogous to any more typical psychometric testing program: It can be specifically developed to fit the needs of individual educational programs and the students they serve. Further, similar to any traditional testing program, the usefulness of DMM procedures is closely related to the relevance of the construct being measured. For example, if trainees' growth rates in medical school are not able to meaningfully inform selection decisions made about those trainees (perhaps because slower growth is not a negative characteristic), then the importance of DMM or any progress-testing paradigm may be limited to formative feedback to students about their improvement. Finally, up until this point in HPE, phrases such as "this student has shown a positive learning trajectory" have sometimes been used as code for educators' desire to place a positive spin on a student's initial learning struggles. But, by applying a dynamic measurement paradigm to HPE, such a description of positive learning trajectory (as indicated by strong growth scores) could be redefined not as a sign of difficulty on the part of the learner but, rather, as high and genuine praise for the way that trainee has improved.

There can be little doubt that, theoretically, a health professions trainee's learning trajectory is an attribute of high importance as those trainees become practicing health professionals. However, no extant testing paradigm of which we are aware within the HPE context is currently capable of estimating a psychometrically reliable quantity that represents learning growth. For this

reason, it is not yet well understood how growth scores relate to the wide variety of learning outcomes associated with HPE, and many future directions for DMM research in the field remain. However, with this proof-of-concept investigation, the promise and potential of DMM to improve the validity of measurement in HPE have been demonstrated, opening the door for future work.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: The study described in this article was approved by the institutional review board at the Uniformed Services University of the Health Sciences.

D. Dumas is assistant professor of research methods and information science, University of Denver, Denver, Colorado.

D. McNeish is assistant professor of quantitative psychology, Arizona State University, Phoenix, Arizona.

D. Schreiber-Gregory is data analyst, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

S.J. Durning is professor of medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

D.M. Torre is associate professor of medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

References

1 Slotnick HB. How doctors learn: Education and learning across the medical-school-to-practice trajectory. Acad Med. 2001;76:1013–1026.
2 McNeish D, Dumas D. Nonlinear growth models as measurement models: A second-order growth curve model for measuring potential. Multivariate Behav Res. 2017;52:61–85.
3 Dumas D, Torre DM, Durning SJ. Using relational reasoning strategies to help improve clinical reasoning practice. Acad Med. 2018;93:709–714.
4 Durning SJ, Costanzo ME, Beckman TJ, et al. Functional neuroimaging correlates of thinking flexibility and knowledge structure in memory: Exploring the relationships between clinical reasoning and diagnostic thinking. Med Teach. 2016;38:570–577.
5 Albanese M, Case SM. Progress testing: Critical analysis and suggested practices. Adv Health Sci Educ Theory Pract. 2016;21:221–234.
6 Schuwirth LW, van der Vleuten CP. The use of progress testing. Perspect Med Educ. 2012;1:24–30.
7 Freeman A, van der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. Med Teach. 2010;32:451–455.
8 Rutgers DR, van Raamt F, van Lankeren W, et al. Fourteen years of progress testing in radiology residency training: Experiences from The Netherlands. Eur Radiol. 2018;28:2208–2215.
9 Pugh D, Touchie C, Humphrey-Murto S, Wood TJ. The OSCE progress test—Measuring clinical skill development over residency training. Med Teach. 2016;38:168–173.
10 Pugh D, Regehr G. Taking the sting out of assessment: Is there a role for progress testing? Med Educ. 2016;50:721–729.
11 Swanson HL. Effects of dynamic testing on the classification of learning disabilities: The predictive and discriminant validity of the Swanson-Cognitive Processing Test. J Psychoeduc Assess. 1995;13:204–229.
12 Calero MD, Belen G-MM, Robles MA. Learning potential in high IQ children: The contribution of dynamic assessment to the identification of gifted children. Learn Individ Differ. 2011;21:176–181.
13 Dumas DG, McNeish DM. Dynamic measurement modeling: Using nonlinear growth models to estimate student learning capacity. Educ Res. 2017;46:284–292.
14 Dumas DG, McNeish DM. Increasing the consequential validity of reading assessment using dynamic measurement modeling. Educ Res. 2018;4:612–614.
15 Harik P, Clauser BE, Grabovsky I, Margolis MJ, Dillon GF, Boulet JR. Relationships among subcomponents of the USMLE Step 2 Clinical Skills Examination, the Step 1, and the Step 2 Clinical Knowledge Examinations. Acad Med. 2006;81(10 suppl):S21–S24.
16 Dong T, Durning SJ, Gilliland WR, Swygert KA, Artino AR Jr. Development and initial validation of a program director's evaluation form for medical school graduates. Mil Med. 2015;180(4 suppl):97–103.
17 Durning SJ, Pangaro LN, Lawrence LL, Waechter D, McManigle J, Jackson JL. The feasibility, reliability, and validity of a program director's (supervisor's) evaluation form for medical school graduates. Acad Med. 2005;80:964–968.
18 McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? Acad Med. 2011;86:48–52.