

Why Does This Learner Perform Poorly on Tests? Using Self-Regulated Learning Theory to Diagnose the Problem and Implement Solutions

Mary A. Andrews, MD, MPH, William F. Kelly, MD, and Kent J. DeZee, MD, MPH

Abstract

Problem

Learners who underperform on standardized tests are common throughout all levels of medical education and require considerable faculty time and effort to remediate. Current methods for remediating test-taking difficulties are typically not grounded in educational theory or supported by high-quality evidence.

Approach

A test-taking assessment was developed based on self-regulated learning (SRL) microanalytic assessment and training and used during academic year 2012–2013. This method assesses the SRL subprocesses

of strategic planning, self-monitoring, causal attributions, and adaptive inferences during the educational task of answering test questions. The method was designed for one-on-one use by teacher and learner, and for learner self-assessment and practice.

Outcomes

At one academic institution, this method was used to categorize learners into struggling test-taker subtypes which correspond to deficiencies in the SRL subprocesses outlined above. A learning plan based on the SRL deficiency was developed for each struggling test-taker subtype. In a small number of internal

medicine residents with low in-training examination scores, use of this method yielded improvements in 2013 in-training examination score that doubled the expected improvement based on historical averages.

Next Steps

This method is a novel application of SRL theory to a commonly encountered problem in medical education: the learner who performs poorly on tests. Large-scale, multicenter studies of medical learners at a variety of training levels and program types are needed to determine the effectiveness and generalizability of this intervention.

Problem

Many medical learners struggle on multiple-choice, clinical-vignette-based knowledge tests, and remediating these learners is resource intensive.¹ Remediation methods range from emphasizing test-taking techniques and completing practice questions, to comprehensive efforts to improve medical knowledge and clinical reasoning.^{1–3} However, recent reviews found only weak evidence of the effectiveness of such interventions.^{4,5}

M.A. Andrews is assistant professor, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

W.F. Kelly is associate professor, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

K.J. DeZee is professor, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

Correspondence should be addressed to Mary A. Andrews, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814-5119; telephone: (301) 295-7879; e-mail: mary.andrews@usuhs.edu.

Acad Med. XXXX;XX:00–00.

First published online

doi: 10.1097/ACM.0000000000001422

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A389>.

Furthermore, remediation methods are often developed and implemented without meaningful basis in educational theory.⁴ Here we introduce a novel, ready-to-use test-taking assessment and remediation framework based on self-regulated learning (SRL) theory; describe implementation at our center; and lay the foundation for a large-scale study of this promising solution to a perennial problem.

SRL theory is an educational assessment framework that has been successfully used to improve performance in many disciplines.^{6,7} SRL has been defined as “self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals.”⁸ SRL subprocesses can be divided into three task-centered categories: forethought, performance, and self-reflection.⁷ Prior research has shown correlations between SRL subprocesses and medical school performance; for example, strategic planning by second-year medical students was significantly associated with United States Medical Licensing Examination Step 1 scores after adjusting for Medical College Admission Test scores.⁹ We postulated that an SRL-based assessment of test-taking

performance would be a useful means of diagnosing and remediating struggling test takers.

Approach

This method was designed as a form of Self-Regulated Learning Microanalytic Assessment and Training (SRL-MAT), which captures the dynamic, context-specific nature of self-regulatory processes by posing questions to the learner at different points during a defined educational task.⁷ The method was developed by one author (K.J.D.) in early 2012, refined by the coauthors, and predates but closely resembles the five-step SRL microanalytic protocol development process described by Cleary et al.⁶ This project was undertaken as a quality improvement initiative and the institutional review board determined that it was not research.

We selected the specific task of answering a multiple-choice, clinical vignette test question of the type commonly encountered on licensing, certification, and in-training examinations (ITEs) and identified target SRL subprocesses for this task: task strategies, self-evaluation, causal attributions, and adaptive inferences.^{7,8} We developed SRL microanalytic questions

specific to the task of answering test questions correctly (Table 1). We designed the task strategy questions to assess the extent to which the trainee uses well-developed disease scripts to arrive at the correct diagnosis, as past experience with struggling test takers and clinical reasoning research have shown this to be an effective strategy.¹⁰ We linked SRL subprocesses to the temporal dimensions of the task by designing the task strategy questions to be administered while the learner works through the test question, self-evaluation questions at specified points during the question–answering process, and causal attribution and adaptive inference questions after the learner was informed whether the selected answer was correct. We developed a categorization system that identified the SRL deficiency based on historical features of the learner’s performance and on the learner’s answers to specific microanalytic questions. This differs from the full transcription and independent coding of open-ended questions recommended by Cleary et al,⁶ but we chose this categorization scheme for practicality and ease of use. This method was designed for one-on-one use by a teacher and learner in a one-hour session, and for learner self-assessment and practice.

The microanalytic questions were combined into a Question Review Form (QRF; see Appendix 1). When

using the form, the teacher first asks the learner to describe prior performance on multiple-choice, clinical-vignette-based medical knowledge tests, including ITEs, steps of licensing examinations, National Board of Medical Examiners subject matter examinations, etc. The teacher then presents the learner with a clinical-vignette-based practice question appropriate to the learner’s training level and asks the learner to engage the question as if it were an actual test question, with the only difference being that the learner should read and think out loud. For this first test question, the learner proceeds through the question without interruption from the teacher or administration of any microanalytic questions. The teacher notes the learner’s level of engagement with the test question, which can range from very superficial (e.g., reading the entirety of the question out loud without any interior commentary or interpretation of the information) to highly engaged (e.g., actively prioritizing and interpreting the clinical information and generating and refining a differential diagnosis while reading the stem). This initial uninterrupted think-aloud exercise may suggest a particular test-taking deficiency, such as underdeveloped disease scripts in the learner who does not readily engage

with the information presented in the clinical vignette.

The teacher then presents a second test question along with the QRF. The teacher covers the last sentence (e.g., “What is the diagnosis?” or “What is the next best step in management?”) and asks the learner to answer QRF Items 1 through 6 after thinking aloud through the clinical vignette. Items 1 through 4 assess the learner’s skill in using the strategy of identifying the disease script in the clinical stem. Items 2 through 4 assess the specificity of the learner’s disease scripts (i.e., how well the learner can recognize different subtypes of disease). Items 5 through 6 are confidence scales which assess the learner’s skill at self-evaluation at defined points in the question–answering process. After completing Item 6, the teacher uncovers the question (e.g., “What is the next best step in management?”) but not the answers. The learner predicts the answer to the question and rates his or her confidence in that answer (Items 7–8). Then the teacher reveals the answer choices and the learner rates his or her confidence in his or her predicted answer (or similar answer) after seeing the choices (Items 9–11), and decides if he or she wants to change his or her answer (Item 12). The teacher then reveals the correct answer, and the learner answers the questions assessing causal attributions and adaptive inferences (Items 14–15). The method is repeated with new test questions until the teacher has accumulated enough evidence to identify one of the test-taking problems outlined in the following section, each of which corresponds to an SRL subprocess deficiency.

As a quality improvement initiative within the internal medicine residency, this assessment tool was administered by one faculty member to 16 second-year internal medicine residents during academic year 2012–2013. A 90-minute workshop, which describes the theoretical basis for the assessment method, reviews how to administer the method, and allows for faculty practice with the method via role-play, was developed as well.

Outcomes

After using this method with approximately 20 struggling learners, we found that most test-taking problems could be categorized as

Table 1

Target SRL Subprocess and SRL Microanalytic Questions for the Task of Answering a Multiple-Choice, Clinical-Vignette-Type Test Question, Uniformed Services University for the Health Sciences, 2012

Target SRL subprocess	SRL microanalytic questions
Task strategies	What diagnosis is the patient most likely to have?
	What is the specific clinical scenario or severity of this disease?
	What factors support your impression of the specific clinical scenario?
	What factors, if any, are inconsistent with your diagnostic/clinical scenario impression?
Metacognitive monitoring and self-evaluation	How confident are you that the patient has the diagnosis you stated above? ^a
	How confident are you in your impression of the specific clinical scenario stated above? ^a
	What is your answer to the question (<i>before looking at the answer choices</i>)?
	How confident are you that your answer will be correct (<i>before looking at the answer choices</i>)? ^a
	How confident are you in your answer now (<i>after looking at the answer choices</i>)? ^a
Causal attributions	Why did you get the answer right or wrong?
Adaptive inferences	What is your plan to improve?

Abbreviation: SRL indicates self-regulated learning.

^aResponse anchors for confidence assessments: not at all confident, slightly confident, moderately confident, quite confident, and extremely confident.

follows: lack of script recognition, lack of script specificity, premature closure, underconfidence, maladaptive causal attributions, and inappropriate adaptive inferences. These test-taking deficiencies are briefly described below; video examples and more details about identifying and remediating each type are available in Supplemental Digital Appendix 1 at <http://links.lww.com/ACADMED/A389>.

The first test-taking deficiency is *lack of script recognition*, which reflects inappropriate or ineffective use of the task strategy of disease scripts for diagnostic reasoning. These learners offer no or factually incorrect interpretations of clinical data while reading the clinical stem, and they cannot answer QRF Items 1 through 4.

The second type of test-taking deficiency is *lack of script specificity*, which is also a task strategy problem. Learners with this deficiency narrow down the answer to two choices by recognizing a general disease category (such as depression), but do not recognize the specific clinical subtype or severity of disease (such as depression in the elderly) and the differences in diagnostic strategy, therapy, or prognosis related to that subtype. These learners will give incomplete or inaccurate answers to QRF Items 2 through 4.

The third test-taking deficiency is *premature closure*, a metacognitive monitoring problem. Learners who exhibit this deficiency make an early decision on the diagnosis, ignore or downplay information inconsistent with the diagnosis, and may list facts that are inconsistent with the chosen diagnosis in QRF Item 4 (yet refuse to change the diagnosis) or simply fail to note the inconsistencies.

The fourth test-taking deficiency is *underconfidence*, which is a metacognitive monitoring and self-evaluation problem. Learners with this deficiency may have a history of several examination failures and have consequently learned to distrust their clinical reasoning, even when it may be correct. During the SRL assessment, these learners rate their confidence on the middle to lower half of the scale (QRF Items 5–6, 8) even when they have identified the correct diagnosis and predicted the correct answer.

The fifth test-taking deficiency is *inappropriate causal attributions*, which

is a deficiency in the self-reflection phase of SRL. This deficiency is evident from superficial or incorrect explanations of why the right answer is correct and the wrong answer is incorrect (QRF Item 14).

The last test-taking deficiency is *inappropriate adaptive inferences*, which is also a self-reflection deficiency. Learners with this deficiency may be able to identify knowledge gaps after answering a practice question, but are unable to articulate an effective learning plan (QRF Item 15).

In using this method to identify test-taking deficiencies and suggest strategies for improvement, we have found improved test performance, albeit in a small number of learners and using historical controls. Over the previous six years in this internal medicine residency, the average second-year ITE score improvement was 4.6 points in the third year. The average improvement in 2013 ITE score from second year to third year for the residents who underwent this assessment was 9.3 points (standard deviation: 7.7), which is about twice the expected improvement based on historical data.

Next Steps

We have introduced a novel application of SRL microanalysis to the common yet challenging problem of learners who underperform on tests.⁷ This method has advantages over existing methods of test-taking remediation, including solid foundation in SRL theory, individualized treatment plans, and relative ease of implementation. Although limited by small numbers, lack of contemporary control group, and the possibility of regression towards the mean, the results of our single-center implementation are promising. While the time needed to train faculty to use this method may vary, we have used a 90-minute workshop format with video examples and role-play (Supplemental Digital Appendix 1, <http://links.lww.com/ACADMED/A389>) at our institution and others.

Recognizing the need for educational interventions to be supported by large-scale studies of feasibility and effectiveness, the next steps for this method are the design and conduct of a large randomized controlled trial.^{4,5} This study should include learners from different programs and training levels, such as students who fail a

subject matter examination or licensing exam and trainees with low ITE scores, and should assess which aspects of the method contribute the most to improvement. While the categorization system outlined above, which relies on historical features plus SRL assessment, has been easily implemented at a single institution, the reliability and validity of this system should be studied in broader settings.

Learners who underperform on multiple-choice, clinical-vignette-based medical knowledge tests are ubiquitous, and poor test performance can have serious professional consequences. Given the paucity of theoretically sound yet ready-to-implement interventions for assisting learners who struggle on tests, the need to disseminate and investigate the effectiveness of this and other interventions is paramount.^{4,5}

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: The Walter Reed National Military Medical Center Department of Research Programs reviewed this project and determined that it was not research.

Disclaimer: The views and opinions expressed in this manuscript are those of the authors and do not necessarily represent the official policy of the United States, Department of Defense, the United States Army, the United States Navy, Walter Reed National Military Medical Center, Tripler Army Medical Center, or Uniformed Services University of the Health Sciences.

Previous presentations: This method has been presented as a workshop at the 2015 Accreditation Council for Graduate Medical Education Annual Educational Conference, February 2015, Orlando, Florida; at the 2015 Clerkship Directors in Internal Medicine Annual Conference, October 2015, Atlanta, Georgia; and at the 2016 Association of Pulmonary and Critical Care Medicine Program Directors Annual Conference, March 2016, San Antonio, Texas.

References

- 1 Guerrasio J, Aagaard EM. Methods and outcomes for the remediation of clinical reasoning. *J Gen Intern Med.* 2014;29:1607–1614.
- 2 Harthun NL, Schirmer BD, Sanfey H. Remediation of low ABSITE scores. *Curr Surg.* 2005;62:539–542.
- 3 Mathis BR, Warm EJ, Schauer DP, Holmboe E, Rouan GW. A multiple choice testing program coupled with a year-long elective experience is associated with improved performance on the internal medicine in-training examination. *J Gen Intern Med.* 2011;26:1253–1257.

4 Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. The remediation challenge: Theoretical and methodological insights from a systematic review. *Med Educ.* 2013;47:242–251.

5 Hauer KE, Ciccone A, Henzel TR, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Acad Med.* 2009;84:1822–1832.

6 Cleary TJ, Callan GL, Zimmerman BJ. Assessing self-regulation as a cyclical, context-specific phenomenon: Overview and analysis of SRL microanalytic protocols. *Educ Res Int.* 2012;2012:19.

7 Durning SJ, Cleary TJ, Sandars J, Hemmer P, Kokotailo P, Artino AR. Perspective: Viewing “strugglers” through a different lens: How a self-regulated learning perspective can help medical educators with assessment and remediation. *Acad Med.* 2011;86:488–495.

8 Zimmerman B. Attaining self-regulation: A social-cognitive perspective. In: Boekarts M, Pintrich P, Zeidner M, eds. *Handbook of Self-Regulation.* Orlando, FL: Academic Press; 2000:13–39.

9 Artino AR Jr, Cleary TJ, Dong T, Hemmer PA, Durning SJ. Exploring clinical reasoning in novices: A self-regulated learning microanalytic assessment approach. *Med Educ.* 2014;48:280–291.

10 Charlin B, Boshuizen HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ.* 2007;41:1178–1184.

Appendix 1

Question Review Form for a Protocol for Test-Taking Assessment Based on Self-Regulated Learning Microanalytic Assessment and Training, Uniformed Services University for the Health Sciences, 2012

Question Review Form

For use with a clinical-vignette-style test question. Cover up the answers and the last sentence of the stem; e.g., the actual question being asked. The learner reads through the stem and answers #1–6 below.

1. What diagnosis is the patient most likely to have? _____
2. What is the specific clinical scenario and/or severity of this disease (for example, if the disease was depression, is this uncomplicated depression, depression in the elderly, depression with history of mania, depression with suicidal ideation, etc.) _____
3. What factor(s) support your impression of the specific clinical scenario? _____
4. What factor(s), if any, are inconsistent with your diagnostic/clinical scenario impression? _____
5. How confident are you that the patient in the test item has the diagnosis you listed in #1 above?

Not at all confident	Slightly confident	Moderately confident	Quite confident	Extremely confident
----------------------	--------------------	----------------------	-----------------	---------------------

6. How confident are you in your impression of the specific clinical scenario in #2 above?

Not at all confident	Slightly confident	Moderately confident	Quite confident	Extremely confident
----------------------	--------------------	----------------------	-----------------	---------------------

Uncover the last sentence of the stem; i.e., the actual question (e.g., “What is the next best step in management?”).

7. Before looking at the answer choices, what is your answer to the question? _____
8. How confident are you that your answer will be correct?

Not at all confident	Slightly confident	Moderately confident	Quite confident	Extremely confident
----------------------	--------------------	----------------------	-----------------	---------------------

Now uncover the answer choices.

9. Does your predicted answer appear? Yes No N/A
10. If no, are there one or more related answers to your predicted answer? Yes No/N/A
11. How confident are you in your answer now?

Not at all confident	Slightly confident	Moderately confident	Quite confident	Extremely confident
----------------------	--------------------	----------------------	-----------------	---------------------

12. If you wish to change your answer (or suggest one if you didn’t predict an answer), list it here: _____

Now look up the answer.

13. Did you get the answer right? Yes No
14. Why or why not? Whether or not you got the item correct, what else do you need to learn? Did you know the exact reason why the right answer was right? Did you know why each fact in the stem was consistent or inconsistent with the clinical scenario? Do you know why the wrong answers are wrong? _____
15. Based on Item 14 (above), what is your plan to improve? _____

The learner should use these forms to collect data on him/herself. How often does the learner get the diagnosis correct? How often does the learner’s clinical scenario match the one in the annotated answer? Is the learner’s confidence in these items appropriate and can he accurately predict the correct answer before looking at the answer choices? How often does changing answers result in the right answer? And so on.