Assessing the Effectiveness of Case-Based Collaborative Learning via Randomized Controlled Trial

5 authors, including:

Edward Krupat
Harvard Medical School
131 PUBLICATIONS   6,644 CITATIONS

SEE PROFILE

Jeremy B Richards
Beth Israel Deaconess Medical Center
89 PUBLICATIONS   610 CITATIONS

SEE PROFILE

Amy M Sullivan
Harvard Medical School
80 PUBLICATIONS   3,005 CITATIONS

SEE PROFILE

Richard Schwartzstein
Beth Israel Deaconess Medical Center
175 PUBLICATIONS   7,005 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project ON TRAC: Improving Asthma Control in Children View project

Ro1 Measurement of Dyspnea View project

Assessing the Effectiveness of Case-Based Collaborative Learning via Randomized Controlled Trial

Edward Krupat, PhD, Jeremy B. Richards, MD, Amy M. Sullivan, EdD, Thomas J. Fleenor Jr, MEd, and Richard M. Schwartzstein, MD

Abstract

Purpose
Case-based collaborative learning (CBCL) is a novel small-group approach that borrows from team-based learning principles and incorporates elements of problem-based learning (PBL) and case-based learning. CBCL includes a preclass readiness assurance process and case-based in-class activities in which students respond to focused, open-ended questions individually, discuss their answers in groups of 4, and then reach consensus in larger groups of 16. This study introduces CBCL and assesses its effectiveness in one course at Harvard Medical School.

Method
In a 2013 randomized controlled trial, 64 medical and dental student volunteers were assigned randomly to one of four 8-person PBL tutorial groups (control; n = 32) or one of two 16-person CBCL tutorial groups (experimental condition; n = 32) as part of a required first-year physiology course. Outcomes for the PBL and CBCL groups were compared using final exam scores, student responses to a postcourse survey, and behavioral coding of portions of video-recorded class sessions.

Results
Overall, the course final exam scores for CBCL and PBL students were not significantly different. However, CBCL students whose mean exam performance in prior courses was below the participant median scored significantly higher than their PBL counterparts on the physiology course final exam. The most common adjectives students used to describe CBCL were "engaging," "fun," and "thought-provoking." Coding of observed behaviors indicated that individual affect was significantly higher in the CBCL groups than in the PBL groups.

Conclusions
CBCL is a viable, engaging, active learning method. It may particularly benefit students with lower academic performance.

This report introduces case-based collaborative learning (CBCL), an alternative small-group instructional method that borrows from the principles of team-based learning (TBL)[1–4] and incorporates elements of both problem-based learning (PBL)[5–7] and case-based learning (CBL).[8,9] Like TBL, CBCL can be used in basic science and clinical instruction, focuses student discussions on faculty-defined topics and learning goals, and aims to promote active learning and high levels of student engagement.

In TBL, students' preclass preparation is encouraged through the readiness assurance process (RAP), whereby each student takes a multiple-choice test in class (the individual readiness assurance test, or I-RAT) based on the topics assigned prior to class. Following the I-RAT, students work in small groups to achieve consensus answers on a group readiness assurance test (G-RAT). Groups' responses are then compared, and a larger group consensus is ultimately reached following discussion. Then, the instructor announces the correct responses, which students may challenge or debate. Michaelsen and Sweet[1] have estimated that the RAP requires approximately 45 to 75 minutes of class time, which may be followed by an additional one to four hours of application-oriented activities using the same format.

In contrast, readiness assurance is accomplished in CBCL in advance of class. Students independently answer challenging multiple-choice questions (MCQs) based on key concepts from the course materials and submit their answers before class. This process encourages students to rely on course materials to generate responses and provides instructors with information about each student's mastery of the material as well as the overall difficulty of the MCQs.

In class, the CBCL instructor presents a new case to students, who generate individual responses to a set of focused, open-ended, case-based questions. Comparing their individual answers, students discuss the alternatives to reach a consensus answer within their assigned small group. The answers of each small group are then shared in the class, and the students try to reach a between-group consensus, with the instructor acting as a facilitator. While this format maintains the general structure of TBL in-class activities, the content of CBCL in-class activities is exclusively case based, as in CBL. In contrast to PBL in which students have greater leeway to explore issues in the case that they deem most relevant, in CBCL the questions answered in class are formulated in advance by the instructor. Because these questions are open-ended, CBCL requires students to generate their own hypotheses and devise creative solutions but, in the spirit of PBL, still encourages independent student exploration of the mechanisms of disease. In these ways, CBCL incorporates some of the best features of PBL, TBL, and CBL.[10]

In this article, we report the results of a randomized controlled trial (RCT) comparing students assigned to CBCL versus traditional PBL groups to assess the impact of CBCL using three different

1

outcomes: final exam performance, student ratings, and coding of observed behavior. We hypothesized that students' learning in CBCL groups would be at least equivalent to that of students in PBL groups (based on final exam performance). We also hypothesized that CBCL students would both be more involved (based on coded observed behavior) and find the process more engaging (based on student survey ratings) than PBL students.

Method

This study was conducted in February–March 2013 in Integrated Human Physiology, a required six-week course for first-year students at Harvard Medical School (HMS). One of the course's central teaching methods is the tutorial, which had always used PBL; all students are required to participate in three tutorial sessions per week. Before the course began in 2013, the course director (R.M.S.) invited students to participate in a test of a new form of tutorial. The first 64 students who volunteered (56 of 132 first-year medical students and 8 of 36 first-year dental students—all of whom take their preclerkship curriculum together) were randomly assigned to a tutorial in the control condition (one of four 8-student PBL tutorial groups; n = 32 students) or in the experimental condition (one of two 16-student CBCL tutorial groups; n = 32 students). All students discussed five cases focusing on different organ systems; each case was covered in three tutorial sessions lasting 70 to 90 minutes each. CBCL and PBL students participated equally in all other aspects of the course (e.g., lecture, patient clinics), took common exams, and received no special instruction other than what was required as part of the course.

Each of the study's six tutorial sections was led by a single tutor. All of these tutors were highly experienced and had previously been rated highly by students. Prior to the course, all tutors participated in a 90-minute faculty development session; in addition, before day 1 of each case, all tutors met with the course director to review learning objectives. The protocol was deemed exempt by the HMS institutional review board.

PBL groups (control)

In PBL tutorials for this course, new elements of a case are introduced over three sessions. Students are encouraged to define the key features of the case via discussion, to seek out new information as needed, and to engage in self-guided explorations that will enable them to explore key basic science concepts in the clinical context and generate explanations for clinical findings according to course learning objectives. The role of the tutor is to monitor the student discussions, contribute as needed, and keep discussions from wandering too far afield. Because we wanted the four PBL control groups of 8 students to be true instances of "standard care," the control group tutors were blinded to the details of the CBCL intervention and operated their tutorials with no reference to being part of an educational experiment.

CBCL groups (experimental)

Readiness assurance. In the CBCL tutorials, each case began on the first day that a new organ system was introduced in the course. The first tutorial session, which introduced the basic elements of the case, began shortly after the initial lecture on the topic, so it was not feasible to expect advance preparation by students on day 1 of a given case. In the second and third tutorial sessions for each case, additional case-based information was presented. Because students already had prior exposure to relevant material, we attempted to ensure readiness by electronically sending them a set of five challenging MCQs the day before the session. The MCQs covered topics relevant to but not specifically about the case, and they required a good deal of application and inference.

Students were instructed to study the course materials and to use the MCQs as a means of assessing their understanding. They were required to work independently and to send their responses to the tutor before the start of the tutorial session. Students were told that failure to submit a response would be considered a breach of professionalism, and scores on each "exercise" contributed an (unspecified) amount to their course tutorial participation grade. After the tutorial session, the correct answers with explanations were posted online for review.

Group discussion. The 32 CBCL students met in two tutorial groups of 16 students. Each student was assigned to the same table of 4 students for the entire course.

The small-group size of 4 was chosen on the basis of social–psychological evidence suggesting that it represents a compromise between smaller groups that may not bring sufficient resources and larger groups in which diffusion of responsibility may take place.[11,12] To the extent possible, we matched the tables of 4 according to gender, society (i.e., students' advisory groups), presence of dental students, and students' overall prior exam performance. (The same procedure was used for assigning students to the study's PBL groups of 8 students.) The CBCL tutors had participated previously in an informal pilot similar to this; their additional training for this study focused on a "script" for conducting CBCL sessions and guidelines for the timing of in-class activities.

On each day of each case, the tutor revealed progressively more information about the patient and condition at the beginning of the session. Immediately after presenting the new information, the tutor provided each student with a sheet containing an open-ended question (see Box 1 for an example). Students were given 5 to 10 minutes to generate individual written responses. Simultaneous reveal of individual responses within the small groups was accomplished by having the students pass their answers to the person on their left; students continued passing until each student had read all four answers. The small groups were then given 15 to 20 minutes to discuss their answers to generate a written consensus response. Simultaneous reveal of the four group consensus responses was accomplished by having each small group post a photo of their response on one of the four large computer screens in the room. This reveal was followed by a 15-minute large-group discussion. The tutor-moderated large-group discussion had the goal of helping students identify the "best" answer (most often one of the four small-group responses, but occasionally a "fusion" of the answers provided) and, more important, the underlying physiological principles supporting the best answers provided. During each session, this procedure was then repeated for a second question about the case.

Outcome measures

We collected three complementary measures: final exam performance as

Box 1

Sample Pulmonary Case Summary and Open-Ended Question for Case-Based Collaborative Learning, With Examples of First-Year Students' Individual In-Class Responses, Integrated Human Physiology Course, Harvard Medical School, 2013

Pulmonary case summary and question

Ms. Liggett is a 53-year-old woman who presents to her primary care doctor with several months of a daily morning cough and progressive shortness of breath with mild exertion. She was last seen 6 years ago for "bronchitis," and her doctor at the time strongly recommended that Ms. Liggett quit smoking. Prior to coming into the examination room, Ms. Liggett performed spirometry, and her forced expiratory volume in one second (FEV1) was 1.6 liters, which is 50% of the predicted value.

Question 1: In light of what you know about the ventilatory pump, provide two physiological hypotheses, based on the patient's history and physical exam, for her increased shortness of breath.

Hypothesis #1:

_____

Hypothesis #2:

_____

Sample responses from students

- Reduced elastic recoil leads to greater compliance which leads to prolonged exhalation
- Restriction of airway due to history of smoking and bronchitis
- Emphysema—decreased elasticity, death of alveolar gas exchange tissue
- Old age leads to loss of elasticity in lung tissue
- Chest muscles do not relax
- Narrowing of airway could decrease resistance
- Decreased chest wall movement from added weight
- Airway inflammation leading to increased resistance and decreased flow
- Decreased muscle tone of chest wall due to inactivity
- Pneumonia—infection
- Fluid/mucous in airways due to heart troubles from smoking

an indicator of *student learning*; student ratings of their tutorial as an indicator of *subjective experience*; and systematic coding of observed class behavior, as captured on video, as an indicator of *process*.

Final exam performance. We compared the standardized scores of all students in the CBCL groups and the PBL groups on a common course final exam. Additionally, on the basis of past findings indicating that TBL may have its largest impact on students who are having academic difficulties,[13,14] we made separate CBCL versus PBL comparisons of the final exam scores of those students whose mean exam performance in prior courses had been above or below the median score for all 64 participating students.

The course final exam was based on two new cases provided to all students for independent study two days before the day of the test. The exam questions, distributed in class on the test day, required

students to explain in short paragraphs the physiological principles underlying the symptoms, physical findings, and laboratory results in the two cases.

Student ratings. Students' perceptions of their tutorial were assessed using a series of ratings on an anonymous postcourse questionnaire hosted at SurveyMonkey (Palo Alto, California). E-mailed invitations with a link to the survey were sent to all study participants shortly after the course ended. The survey items were written to assess subjective experience along three dimensions: learning effectiveness (12 items such as "Tutorial was challenging and thought-provoking" and "Tutorial enhanced my critical thinking skills"; alpha = 0.92); positive affect (8 items such as "I felt excited and enthusiastic during discussions" and "I looked forward to coming to tutorial"; alpha = 0.76); and preparedness (2 items: "I felt that students were quite well prepared when they came to tutorial" and "I typically put a great deal of effort preparing for tutorial in advance";

alpha = 0.67). All items were answered using the same four-point Likert scale (from agree = 1 to disagree = 4), analyzed so that a rating of 1 always represented the most positive response. In addition, a free-text item asked students to list two adjectives that best described their tutorial.

Coding of observed behavior. Fourteen tutorial sessions were video-recorded for each PBL and CBCL group (five cases × three sessions, minus one session because of technical problems). In the PBL sessions, one camera was used to capture picture and sound for all 8 students. In the CBCL sessions, four wide-angle cameras were used, with each camera focused on a single table of 4 students. Because four small-group discussions were occurring simultaneously in the CBCL sessions, separate voice recordings of each table's discussions were collected using strategically placed, directionally sensitive microphones, and voice and video were synced. During the 16-person CBCL discussions, we used only one of the four cameras, which was located behind the tutor and therefore capable of capturing picture and sound for the larger-group discussions.

Behavioral coding was accomplished using the thin slices method,[15–17] a validated social science approach, and all behavioral ratings were made by two trained coders. From within the specified discussion periods, we randomly selected 30-second video snippets (or "slices") of group interaction for observation. For each recorded PBL session, we selected eight slices: two from early in the session, four in the middle, and two late in the session. For each recorded CBCL session, we selected eight slices during each 4-person group discussion (distributed similarly to the PBL slices) and another eight slices during the 16-person large-group discussion.

Each slice was coded several ways, adapting an approach that has been validated in other studies.[18,19] Group affect was rated on six dimensions (tense, warm, cooperative, engaged, energetic, open) using a five-point scale (ranging from a score of 1 to indicate that the characteristic was definitely present to 5 to indicate definitely absent); these ratings were summed into a single index. Individual affect was coded by focusing on the behavior of a randomly selected individual student. The coders noted

whether that person was primarily passive (i.e., listening) or active (i.e., talking to another student or the tutor) during the slice, and they rated that person's affect on six dimensions (enthusiastic, cheerful, active, tense, engaged, energetic) using the same five-point scale as above; these ratings were also summed into a single index. Finally, coders noted the presence or absence of eight types of individual behavior (e.g., giving information, agreeing, expressing reassurance) during all of the tutorial session group discussions.

To avoid halo effects, the coders were instructed to focus on specific, single behaviors. To eliminate any demand effects or rater bias, the coders were told that we were not attempting to establish the superiority of one method or the other. With one of the authors (E.K.), the coders discussed the conceptual distinctions among the various characteristics to be observed. They then selected several slices, made independent ratings, and discussed differences in ratings when they occurred. This process continued until they believed that they had reached a high level of agreement, at which point they calculated Cronbach's alpha based on their most recent ratings. When the interrater consistency reached the criterion of alpha = 0.7, the coders began coding independently. At the midpoint of coding, rater drift was assessed. Interrater reliability was still above the criterion, and the coders returned to independent coding.

We were able to code 402 slices of the PBL discussions and an additional 400 to 410 slices of the CBCL 4-person and 16-person discussions, for a total of over 800 CBCL observations. (The variability resulted from technical problems including issues related to making uniform behavioral observations in the CBCL rooms where four cameras were used.) After several practice sessions, interrater reliability reached a satisfactory level (alpha = 0.79 for individual affect, 0.79 for group affect, and 0.85 for group behavior). Internal consistency of the group affect and individual affect indices were high (alpha = 0.74 and 0.97, respectively).

Statistical analysis

Our study design was nested, with 32 students in the CBCL group and 32 in the PBL group. As each group member could not be considered an

independent observation, we used a multilevel modeling approach in which statistical comparisons between the CBCL and PBL groups were made using univariate between-subjects t tests or analysis of variance (ANOVA) with nested interaction terms. Analyses were conducted using SPSS version 21.0 (IBM SPSS, Armonk, New York).

Results

Final exam performance

The standardized final exam scores for the CBCL students and the PBL students (53.50 versus 47.13, respectively) were not significantly different ($t = 0.92$; $P = .36$). However, the mean final exam score of the CBCL students whose mean exam scores in prior courses were below the median of the 64 participants was significantly higher than that of their PBL counterparts (41.63 versus 26.88, respectively; $t = 2.04$; $P = .05$). There was no significant difference in the mean final exam scores of CBCL and PBL students whose mean prior exam scores were above the median (see Figure 1).
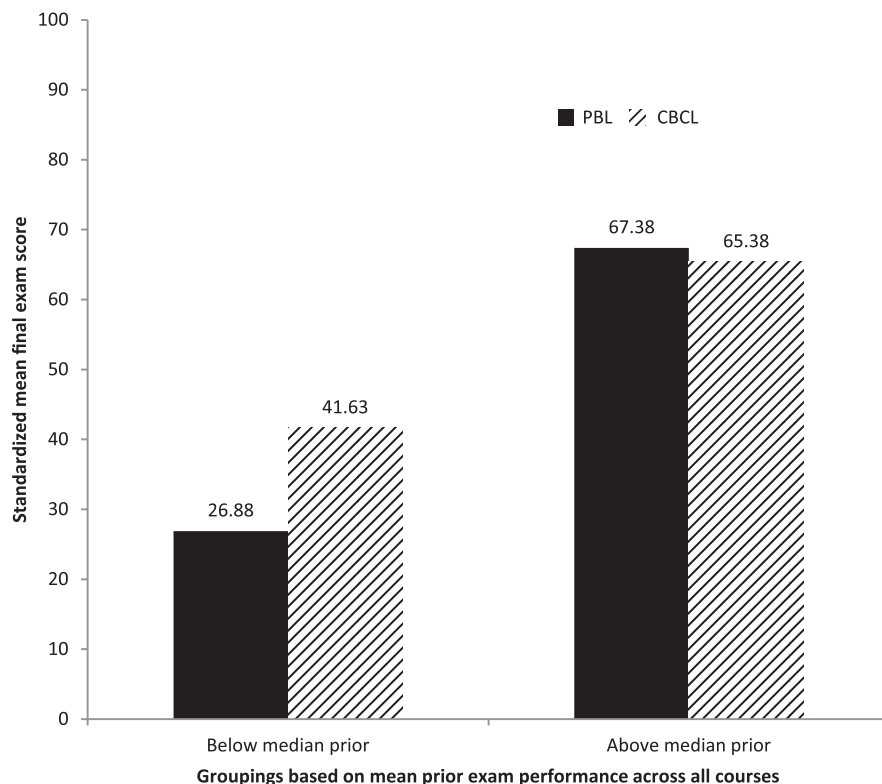
Student ratings

The response rates for the postcourse survey were 91% (29/32) in the PBL group and 97% (31/32) in the CBCL group. CBCL students' ratings on all three dimensions were consistently more positive than those of PBL students, but none of the differences reached statistical significance. The majority of free-text responses were positive in both the PBL and CBCL groups; however, 12% of the PBL terms were negative (e.g., "boring," "frustrating," "scattered"), whereas only one term for CBCL could be described as negative ("indecisive"). As illustrated in Figure 2, the words most used in the PBL group's free-text responses were "fun" (n = 6) and "collaborative" (n = 4). In the CBCL group's responses, the most common words were "engaging" (n = 9), "fun" (n = 6), and "thought-provoking" (n = 5).

Coding of observed behavior

On the combined index of individual affect, students were significantly more positive in the CBCL groups than in the PBL groups ($P < .0002$). There were no significant differences between the PBL and CBCL groups on group affect. The analysis

Figure 1 Comparison of standardized mean scores on the March 2013 Integrated Human Physiology final exam, Harvard Medical School, for first-year students whose mean prior exam performance across all courses was above or below the median for all 64 students participating in the study. PBL indicates students assigned to the study's problem-based learning tutorial groups (control condition); CBCL, students assigned to the study's case-based collaborative learning tutorial groups (experimental condition).

Figure 2 Word clouds representing first-year students' free-text responses when asked for two adjectives that best describe their tutorial experience as part of the Integrated Human Physiology course, Harvard Medical School, postcourse survey, March 2013. Panel A depicts responses from 29 of the 32 students assigned to the study's problem-based learning groups (control condition). Panel B depicts responses from 31 of the 32 students assigned to the study's case-based collaborative learning groups (experimental condition). The physical size of each word indicates the relative frequency of mentions by the respondents.

of specific behaviors (i.e., presence versus absence per slice) revealed a complex pattern: There were fewer explicit instances of asking for or giving information in the CBCL small-group discussions than in the PBL discussions, as well as fewer expressions of uncertainty, respectful disagreement, and reassurance; yet there was a higher incidence of both expression of frustration and lightheartedness (see Table 1). During any given 30-second slice, the student randomly chosen for observation was primarily active (speaking or engaged in a form of interaction) 42% of the time in a group of 4, 14% of the time in a group of 8, and 6% of the time in a group of 16 ($P < .0001$).

Discussion

This RCT compared CBCL, a new team-based small-group approach, versus PBL, a time-tested and widely accepted tutorial model. While performance on the course final exam was not significantly higher for CBCL students than for PBL students overall, those CBCL students whose prior exam performance was below the median of all study participants had a significantly higher mean score on the course final exam than their PBL counterparts. This finding underscores the value of CBCL as a means of enhancing the performance of students about whom educators have the most concern.

This result is noteworthy given three factors that work against the likelihood of finding performance differences between the CBCL and PBL groups. First, the final exam was based at least as much on common course elements (e.g., readings, lectures, other teaching modalities) as it was on the tutorials, yet CBCL–PBL performance differences were still found. Second, given the common assertion that the tutor's ability can make the

critical difference in learning regardless of pedagogical approach, we selected only the most highly rated tutors for this study to control for this potentially confounding variable, yet we still found a CBCL–PBL difference. Third, the CBCL tutors were working with a new and somewhat more complicated teaching format, which could have adversely affected student learning but apparently did not.

Similar to the final exam scores, students' ratings of their tutorial experience on the postcourse survey were generally, but not significantly, higher in the CBCL group than the PBL group. This may be due to ceiling effects, as students were favorably disposed to both teaching methods. Although it was not surprising that both CBCL and PBL students' free-text responses were primarily positive, the most frequent term CBCL students spontaneously mentioned—"engaging"—matched our central objective for this method. The other terms CBCL students mentioned most frequently—"fun" and "thought-provoking"—were also consistent with our goals.

Behavioral coding indicated that individual affect was significantly more positive in the CBCL students than in the PBL students and that, as would be expected, students were more active in groups of 4 than in groups of 8 and more active in groups of 8 than in groups of 16. The coding of specific behaviors produced a complex picture, with students in CBCL small groups making explicit requests for information and giving information less often than students in PBL groups, even though the final exam results suggest that a great deal of information exchange likely took place in the CBCL small-group discussions.

In considering the relationship of CBCL to TBL (the method whose principles and general structure served as its inspiration), we recognize that both CBCL and TBL share a commitment to the assurance of student readiness. However, the differences between CBCL and TBL raise some important questions that suggest the need for a deeper exploration as to "what is readiness?" and "readiness for what?" CBCL aims to ensure that students are ready to discuss case-based material in class by requiring them to submit answers to a set of challenging

Table 1

Behaviors Observed Among First-Year Students in "Slices" of Video-Recorded Tutorial Discussions, by Tutorial Type, Integrated Human Physiology Course, Harvard Medical School, 2013[a]

Behavior	CBCL discussion slices, no. (%)[b]	PBL discussion slices, no. (% of 402)	z score	P value
Asking for information	274/409 (67)	314 (78)	3.51	.001
Giving information	384/408 (94)	394 (98)	2.90	.002
Expression of respectful disagreement	41/410 (10)	64 (16)	2.44	.007
Expression of agreement	278/409 (68)	285 (71)	0.93	.18
Expression of uncertainty	180/409 (44)	209 (54)	2.85	.002
Expression of frustration	16/400 (4)	4 (1)	2.50	.006
Expression of reassurance	180/409 (44)	209 (52)	2.28	.01
Lightheartedness	119/410 (29)	96 (24)	1.61	.05

Abbreviations: CBCL indicates case-based collaborative learning; PBL, problem-based learning.

[a]Thirty-second "slices" of group interactions during tutorial sessions were randomly selected from 14 systematically selected video-recorded tutorial sessions (see Method for details). All ratings of behavior were made by two trained observers.

[b]The number of slices in which each behavior could be rated varied because of technical difficulties. All CBCL slices included in this table are slices of the 4-person, small-group discussions; slices of the 16-person large-group discussions were not included in this analysis.

MCQs prior to class; this also serves the purpose of focusing their preparation. In TBL, assurance rests on the assumption that students who know they will be tested in the classroom will prepare before coming to class, at which point class time will be devoted to the I-RAT and G-RAT process to ensure their readiness to relate basic science course material to clinical problems. While both approaches are likely to generate prepared and knowledgeable students, CBCL focuses class time exclusively on case-based problems and clinical applications, which may be particularly useful when class time is limited and/or when clinical application is emphasized.

As for the use of MCQs versus open-ended questions for in-class discussions, many colleagues with TBL experience urged us to avoid using open-ended questions in CBCL because student responses would be so variable that small-group discussions would be unfocused. With this in mind, we attempted to craft open-ended questions that were sufficiently focused and case based. Most important, we required all CBCL students to commit to a written answer and subsequently reveal that answer, exactly as written, to the other members of their small groups. With four specific and discrete response alternatives before each small group, the four-person discussions were generally focused and

the groups were capable of generating a specific consensus-based answer. Positive results using open-ended questions for in-class discussions in TBL have recently been reported by others.[20]

Asking students to respond to open-ended questions in the manner described above has certain advantages that we believe are consistent with the kinds of clinical reasoning processes most educators desire to inculcate in their students. First, an open-ended question requiring students to generate their own response alternatives involves what learning psychologists refer to as a "recall task," which is cognitively more complex than a "recognition task" such as an MCQ requiring students to choose among preselected alternatives.[21,22] Second, responding to open-ended questions more closely simulates a clinical task, as a set of four alternatives does not automatically appear after a patient's presentation.

Third, reviewing responses to open-ended questions offers instructors insight into students' reasoning in ways that answers to MCQs do not. In this study, the content and quality of the answers to open-ended questions that CBCL students produced ranged widely in quality, type, and specificity (see bottom panel of Box 1 for examples). The discussions in the 4-person groups sometimes varied considerably because each small group considered and

chose a "best" response from among a range of self-generated alternatives. This also generated 16-person group discussions that were rich with teaching opportunities, each unique to the ideas that group members brought to bear on the question. This variety of responses, however, requires tutors to be able to improvise because they cannot anticipate the exact set of answers that students may generate for consideration. Finally, responding to open-ended questions allows students to gain a better appreciation for the ambiguity of clinical medicine by recognizing that sometimes the answers they seek are not as black and white as desired. This is contrary to what may be suggested by MCQs, even if the process of searching for a "best" rather than "correct" answer may frustrate students at times.

Although this first attempt at CBCL has been successful, we recognize that the findings presented here have limitations. We introduced and evaluated the CBCL method in a single required course for first-year medical students at a single medical school, and the sample size was small. Further tests of CBCL's effectiveness are necessary before any general conclusions about it can be reached.

Nonetheless, we are optimistic about CBCL's potential, given the lessons we have learned that will help us to further improve this method. For example, simultaneous reveal of students' responses to the open-ended questions is essential to make the small-group discussions productive, and we struggled initially with how best to accomplish this. Judging the length of time required for the discussions has been a work in progress, and we have begun to establish better guidelines regarding the time needed for each activity and how much discretion tutors should have to improvise. We are also looking for ways to satisfy our students' desire to know in absolute terms the "correct answer," while encouraging them to focus on the principles that make some answers better than others.

Conclusions

In this first test of CBCL via RCT, we used several different measures to evaluate this new approach. We recognize that the conclusions that can be drawn from this single educational experiment are limited. However, on the basis of our

initial results, we offer CBCL as an alternative to traditional PBL and TBL approaches—which is consistent with the call[10] to find means of combining some of the best features of both. Our data suggest possible benefits deriving from CBCL not only in the learning experience but also in performance, especially for those students who are having the most difficulty mastering the curriculum.

E. Krupat is director, Center for Evaluation, Harvard Medical School, and associate professor of psychology, Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, Massachusetts.

J.B. Richards is assistant professor, Department of Medicine, Medical University of South Carolina, Charleston, South Carolina. At the time of the study, J.B.R. was assistant professor of medicine, Harvard Medical School, Boston, Massachusetts.

A.M. Sullivan is director for research, The Academy at Harvard Medical School, and assistant professor, Department of Medicine, Harvard Medical School, and director for education research, Shapiro Institute for Education and Research, Beth Israel Deaconess Medical Center, Boston, Massachusetts.

T.J. Fleenor Jr is project coordinator, Center for Evaluation, Harvard Medical School, Boston, Massachusetts.

R.M. Schwartzstein is Ellen and Melvin Gordon Professor of Medicine and Medical Education and faculty associate dean for medical education, Harvard Medical School, vice president for education, Beth Israel Deaconess Medical Center, and executive director, Carl J. Shapiro Institute for Education and Research, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, Massachusetts.

References

1 Michaelsen LK, Sweet M. The essential elements of team-based learning. New Directions Teach Learn. 2008;116:7–27.

2 Burgess AW, McGregor DM, Mellis CM. Applying established guidelines to team-based learning programs in medical schools: A systematic review. Acad Med. 2014;89:678–688.

3 Haidet P, Levine RE, Parmelee DX, et al. Perspective: Guidelines for reporting team-based learning activities in the medical and health sciences education literature. Acad Med. 2012;87:292–299.

4 Sisk RJ. Team-based learning: Systematic research review. J Nurs Educ. 2011;50:665–669.

5 Pluta WJ, Richards BF, Mutnick A. PBL and beyond: Trends in collaborative learning. Teach Learn Med. 2013;25(suppl 1):S9–S16.

6 Kinkade S. A snapshot of the status of problem-based learning in U. S. medical schools, 2003–04. Acad Med. 2005;80:300–301.

7 Taylor D, Miflin B. Problem-based learning: Where are we now? Med Teach. 2008;30:742–763.

8 Christensen CM, Carlile PR. Course research: Using the case method to build and teach management theory. Acad Manage Learn Educ. 2009;8:240–251.

9 Avery J. Leveraging crowdsourced peer-to-peer assessments to enhance the case method of learning. J Adv Marketing Educ. 2014;22:1–15.

10 Dolmans D, Michaelsen L, van Merriënboer J, van der Vleuten C. Should we choose between problem-based learning and team-based learning? No, combine the best of both worlds! Med Teach. 2015;37:354–359.

11 Wheelan SA. Group size, group development, and group productivity. Small Group Res. 2009;40:247–262.

12 Hackman JR, Vidmar N. Effects of size and task type on group performance and member reactions. Sociometry. 1970;33:37–54.

13 Koles PG, Stolfi A, Borges NJ, Nelson S, Parmelee DX. The impact of team-based learning on medical students' academic performance. Acad Med. 2010;85:1739–1745.

14 Koles P, Nelson S, Stolfi A, Parmelee D, Destephen D. Active learning in a Year 2 pathology curriculum. Med Educ. 2005;39:1045–1055.

15 Wood TJ. Exploring the role of first impressions in rater-based assessments. Adv Health Sci Educ Theory Pract. 2014;19:409–427.

16 Ambady N, Rosenthal R. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. J Pers Soc Psychol. 1993;64:431–441.

17 Ambady N, Laplante D, Nguyen T, Rosenthal R, Chaumeton N, Levinson W. Surgeons' tone of voice: A clue to malpractice history. Surgery. 2002;132:5–9.

18 Roter DL, Hall JA, Blanch-Hartigan D, Larson S, Frankel RM. Slicing it thin: New methods for brief sampling analysis using RIAS-coded medical dialogue. Patient Educ Couns. 2011;82:410–419.

19 Hall JA, Roter DL, Blanch DC, Frankel RM. Observer-rated rapport in interactions between medical students and standardized patients. Patient Educ Couns. 2009;76:323–327.

20 Badgett RG, Stone J, Collins TC. The importance of free-text responses in team-based learning design. Acad Med. 2014;89:1578.

21 Hall JW, Grossman LR, Elwood KD. Differences in encoding for free recall vs. recognition. Mem Cognit. 1976;4:507–513.

22 Anderson JR, Bower GH. Recognition and retrieval processes in free recall. Psychol Rev. 1972;79:97–123.