Consider a multiple-choice examination in a high-school or college course. Students shuffle in and are seated in a crowded classroom. The instructor, who has made 90 collated, stapled copies of the test, first distributes either fill-in-the-bubble or plain-numbered answer sheets, followed by the test. The instructor tries to prevent cheating attempts as some students are tempted to look at their neighbor’s answer sheet and/or slip an extra copy of the test in their personal items to give to their friends later. Once the allotted time expires, students turn in their tests and answer sheets and leave the classroom, with the assumption that they will be informed of the results in, perhaps, a few days to a week. The instructor then gathers the answer sheets and scores them or takes the fill-in-the-bubble sheets to a testing center to be scored.
As students, each of us remembers well the drained-but-anxious emotional state in which we would leave the classroom after a test. Though tired, we were fueled by a strong desire to know the correct answers to questions on the test and get a sense for our performance. As teachers today, we are reminded of this experience as we pass students in the hallway after a test, going through their books and notes, talking about particular items, and trying to get closure on the testing experience.

But life goes on, and after an hour or so this emotional energy drains away and when students are informed days later of their performance, their reaction is less about understanding why “a,” instead of “d,” was correct for #13, and more about how they did. In short, the incredible learning opportunity that is afforded by students’ emotional energy after taking a test is completely lost. The importance of feedback contiguity has a long lineage in educational research (for a review, see Mory, 2003). Also, feedback is more informative when one can bring the conditions that caused the behavior to memory (Anderson & Lebiere, 1998).

The following week when instructors hand back the scored tests they may even devote class time to going over the test and the results. At this time, a mob mentality may set in where the students grab their torches and pitchforks when they realize that several of their compatriots missed the same “unfair” item. Collective anger is directed at the instructor for including such a “tricky” item and the students demand retribution in the form of discarding the item. The battle lines are drawn with the instructor on the defensive as the anger on both sides escalates. If the instructor gives in to appease the students’ feelings, student respect for the instructor is lost. If the instructor refuses to budge, however, student resentment grows as a cancer, harming any chance of community in the classroom. Either way, an “us versus them” mentality can set in and last for a dishearteningly long time. Of course, instructors may choose to simply not review their tests at all, with the unfortunate consequence of robbing their tests of any instructional value, reducing testing to a purely evaluative (or even punitive) experience.

So, let us review four obvious problems with the typical testing session. First, in crowded classrooms, there is a possibility of cheating by looking at another’s answers and by taking an extra copy of the test from the classroom. Second, the instructor has to make paper copies of the exam, score the answer sheets by hand or take fill-in-the-bubble sheets to a testing center, both consuming instructor time. Third, the students do not receive immediate feedback regarding their performance, thus wasting emotional energy that could be used to more fully internalize the material. Fourth, student anger and frustration regarding certain items they missed are directed solely at the instructor. In this chapter, we present a solution to these problems.
TEAM-BASED TESTING

For the past several years, we have tried to use testing not only as an assessment opportunity but also as a learning opportunity by having students take the same test in teams following the individual test. In the last ten years, collaborative testing has been explored in different forms, across several disciplines (e.g., Cortright, Collins, Rodenbaugh & DiCarlo, 2003; Lusk & Conklin, 2003; Mitchell & Melton, 2003; Zimbardo et al., 2003; Lambiotte et al., 1987). Because it is the most well-documented and widely-used version of this practice, we have modeled our team-based testing after the Readiness Assurance Process (RAP) used in Team-based Learning (Michaelsen, Knight & Fink, 2004). At the beginning of the semester, teams are formed using stratified random sampling to ensure that student characteristics that may influence team performance (e.g., previous coursework) or possibly lead to the formation of factions within teams (gender, major) are distributed evenly across teams. A RAP involves students first taking a short, multiple-choice test over the assigned readings, frequently called an individual Readiness Assurance Test (iRAT). After students turn in the completed iRATs, they then take the exact same test as a team (frequently called a group Readiness Assurance Test, or gRAT), engaging in whatever discussion is necessary to reach consensus on answers for each question. For this process, Michaelsen, Knight, and Fink (2004) strongly recommend using self-scoring answer sheets such as the Immediate Feedback-Answer Technique (IF-AT®) scratch-off forms—a multiple-choice assessment and learning system developed by Epstein Educational Enterprises (epsteineducation.com) (see Figure 11.1).

With the self-scoring answer sheets, students scratch off an opaque film covering an answer box in search of the mark indicating a correct answer (a star). As a result, teams receive truly immediate feedback with every choice they make. If the team finds a star on the first try, they receive full credit. If not, they continue scratching until they do find the star, but their score is reduced with each successive scratch, with partial credit motivating continued discussion and learning (see Figure 11.2).

For any given item on the gRAT, team discussion usually begins with team-members soliciting input from each other about how they answered the item individually. In the face of disagreement, students then ask each other to justify their answers. In those cases where logic fails to persuade group members into consensus, students then resort to a fascinating set of strategies to communicate the various ways they feel about a certain answer. For example, we often hear students say such things as, “I chose B, but I just guessed,” “It’s C—I’m positive.” or “I’ll bet you lunch that it is D.” These statements refer to the level of confidence students have in their answers. They may also refer to their or other team members’ credibility: “Don’t
Figure 11.1  Example of a five-option IF-AT®.

Figure 11.2  Example of an IF-AT® with scores for successful 1st and 2nd attempts.
listen to me—I’ve gotten the last four wrong,” or “You usually get this part right, so let’s go with your answer.” These latter statements especially surface when the team’s first attempt is wrong and they are deliberating on the second attempt.

The team deliberation process corrects one of the problems identified earlier with the individual testing situation. When a peer/teammate explains why a particular choice is incorrect and why another is correct, a student may be less likely to feel that the item was “unfair” and direct anger toward the instructor. Also, because the teams typically perform very well on the test, students are less likely to feel that the test was too difficult. More importantly, students whose understanding of the content is less than complete receive explanations from their teammates during the deliberation process enabling them to gain a better understanding of the content than if they were allowed to simply leave the classroom after the individual test.

Besides the team deliberation process, there is one other mechanism within the RAP that facilitates learning by exploiting the emotional charge students feel immediately after testing. When teams complete their gRAT, they can appeal those missed questions for which they feel they can make a convincing case that: (a) the answer they chose can be considered correct, (b) the question needs to be revised, or (c) the reading material was inadequate. Only teams, not individuals, can submit appeals, and they must be supported by evidence from the course materials (readings, lectures, etc.). This appeal process is designed to drive the students immediately back into the areas of the content they did not understand in pursuit of either clarification or support for their appeal. In our experience, teams often learn a tremendous amount through the process of trying to decide the grounds and wording of an appeal—as an instructor, it is an exciting process to overhear.

Instructors can either review and address team appeals immediately, or review them after class and report later which appeals they have decided to grant. The latter is recommended, based upon the emotional energy generated by the RAP. In either case, it is important for instructors to explain their rationale for appeal decisions, as this keeps the discussion instructional and content-oriented. If an appeal is granted, only those teams that successfully appealed the item receive points—teams cannot “piggyback” on appeals written by other teams.

Once the appeal process is over, the RAP is complete. The instructor can now review for the class those questions that most teams missed on their gRATs and briefly clarify the concepts related to these questions. Questions which all groups answered correctly can be considered as representing ma-
terial already mastered by the process of individual preparation and group discussion. This relieves the instructor of the burden of lecturing over content that students are able to understand on their own, thus freeing up a great deal of class time in which students can learn to put course material to use in application oriented activities.

The RAP, then, serves three purposes. First, it motivates students to read and study the required readings before discussing them in class. Second, it allows student misconceptions and misunderstandings to be clarified by either team members through discussion during the gRAT or by the instructor using the gRAT diagnostically to identify areas of misunderstanding. Finally, the immediate feedback after gRAT discussions tends to develop groups into increasingly cohesive and effective learning teams. Fink (2004) maintained that

> when there is a delay, even a day or two, between when the groups do their work and when the assessment comes back, the typical reaction is to “see what they got” and move on. By contrast, when the teams receive immediate feedback on individual and team performance, they instinctively and inevitably engage in an analysis of what went wrong when they still have a clear enough recollection of their experience to make the necessary corrections…. Thus, in most cases, providing immediate feedback is all that is required to quickly and effectively improve the quality of their learning and performance. (pp. 17–18)

Team-based Learning is a powerful and comprehensive way to teach, and Michaelsen and his colleagues have done a great deal of work to integrate the RAP into its overall instructional strategy (Michaelsen, Knight & Fink, 2004). In fact, the “Readiness” in Readiness Assurance Process refers to the students’ readiness to move on into higher-level, application-oriented activities as prescribed by Team-based Learning.

We have found Team-based Learning much easier to implement in courses where application of concepts and sequential instruction are more obvious (e.g., statistics) than for other courses where the material is designed to be more conceptual and introductory (e.g., human learning). Regardless of whether the other elements of Team-based Learning are in place within a course, we feel this testing sequence is an excellent practice in its own right and have used it in multiple choice testing situations as part of the typical tests or quizzes that follow instructional units. We therefore describe this sequence for the rest of this chapter not in terms of a team’s “readiness” to go on to the next Team-based Learning activity, but simply in terms of its own features: team-based testing (TBT).
For the first few years, we used the IF-AT® forms for TBT. Although we, and especially the students, enjoyed the IF-AT® forms, they do not solve all the problems associated with paper-based testing mentioned earlier. First, distributing paper versions of the tests compromised test and item security. We tried to be diligent in ensuring that each student received only one test but sometimes the stapled copies would stick together. Also, sometimes students would leave without turning in the tests. Second, typically the course is full and students are seated next to each other in close quarters when taking the tests. The opportunity for peeking at others’ answers was almost tempting. Third, the individual and team tests always had to be hand-scored which took a substantial amount of time. Fourth, the IF-ATs® have only one answer key per 250 sheets, meaning that there is a possibility that students may figure out the key if several tests containing several items are given.

In this chapter we introduce an online testing program called team-based testing (TBT) that was developed by the authors and the Learning Technology Center at the University of Texas. TBT was designed to mimic the IF-AT® by allowing students to enjoy all the advantages while at the same time overcoming some of the weaknesses associated with the IF-AT® by protecting test security, making it more difficult to copy answers from another student, and reducing the time needed for scoring and entering grades.

There are numerous online testing programs out there and perhaps one that which most postsecondary instructors are familiar is a feature within an online course management system (e.g., Blackboard, WebCT). We certainly considered simply using the testing feature in these management systems to avoid reinventing the wheel, so to speak. However, there were a few problems with the Blackboard testing feature that interfered with the team test portion of TBT afforded by the IF-AT®. First, Blackboard requires that an individual student log in and take the team test for the team. That does not seem to be problematic until you consider how the other team members will be able to view the test questions. Huddling around a desktop or laptop computer discourages face-to-face dialogue and eye contact—considered essential for the cohesion building that occurs when teammates are learning about each other’s confidence and credibility. To get around this, one could simply provide paper copies of the test to the other teammates. Unfortunately, this brings us back to the test security problem of paper-and-pencil testing. Allowing students to simply review their individual test is another possible solution. Unfortunately, Blackboard also allows them to view their answers and the correct answer. We tried simply preventing students from viewing their individual test results but Blackboard also records their test score in the grade book. The deliberation process during team testing changes dramatically when teams learn that one of their teammates re-
ceived a perfect score. Second, individuals need to know how they answered for each question. Sure, this is not difficult to remember when the test is 20 items or less. But with more items, students’ memory for how they answered becomes weaker. Finally, teams need immediate feedback after each question—similar to using the IF-AT® forms. Blackboard currently does not allow for immediate item feedback. Thus, we needed to develop an online testing system that would (a) not allow students to see correct answers or test scores after the individual test, (b) allow students to view their choices for questions on the individual test, and (c) provide immediate item feedback on the team test, with point reductions for multiple attempts.

**HOW COMPUTER-BASED TBT WORKS**

Students first take the individual test in a computer lab setting with laptops wirelessly connected to the campus network. Why laptops and not bolted-down desktop computers? The laptops allow for mobility and seating with no power or tables (students may actually place the laptops on their laps) and are ideal because they enable students to face each other while they deliberate. Kawashima et al. (1999) described the power of social eye contact to trigger the emotional centers of the brain, and we have seen student discussions become much more intense when they are engaged face to face, as opposed to crowding around—a single desktop computer instead of their teammates.

Students open a web browser and log into the TBT system to take the individual test. Both the individual and team tests are password protected and the instructors give the students the passwords immediately before the test begins. Items on the individual test are presented in a random order, one-at-a-time with no opportunity to go back and change answers. We use this format to prevent students from seeing the same item on their neighbor’s laptop or seeing one and then going back and changing their answer. Figure 11.3 displays a typical item view for students on the individual test. Once students are finished with the individual test, they wait for their teammates to assemble around a table with them and then they log in to take the team test. The system requires them to select a team leader who submits answers. Figure 11.4 displays the interface used to select participants and a test leader. This latter person is the only participant allowed to see the immediate feedback for each item. The test leader communicates the feedback to teammates after each item.

For each item, team members are allowed to see the item and their choice for the item on the individual RAT (Figure 11.5). After discussion, the team leader submits their choice and the feedback is either “correct” or “try again.” If they miss the item, the teams are allowed to submit a second choice with their first incorrect choice grayed out (Figure 11.6). If they an-
15 of 30 questions

15. Marietta does her homework primarily to gain her teacher’s approval. She has little internal desire to do her homework, although she does feel slight twinges of guilt when she fails to get an assignment done on time. Of the four stages through which internalized motivation evolves, Marietta appears to be in the _______ stage.

- external regulation
- identification
- integration
- internalization
- injection

Submit

Figure 11.3 Example of an item view on the individual test.

![Figure 11.3 Example of an item view on the individual test.](image)

**Figure 11.3** Example of an item view on the individual test.

![Figure 11.4 Interface for selecting team participants and leader before taking the team test.](image)

**Figure 11.4** Interface for selecting team participants and leader before taking the team test.

![Figure 11.5 Example of item view on the team test.](image)

**Figure 11.5** Example of item view on the team test.

![Figure 11.6 Example of feedback after one failed attempt.](image)

**Figure 11.6** Example of feedback after one failed attempt.
swer correctly, they get “correct” and an updated point total showing that they received half-credit for that item; whereas, if they answer incorrectly, they get “wrong” and no points for that item. This deduction of points for each failed attempt is a feature we have always used with the IF-AT® forms. It punishes teams for wrong choices but also allows them to get partial credit for demonstrating knowledge with successful successive choices. Once the team submits their final choice, students may review both their team and individual tests online.

As a final benefit—and this is no small matter—all students’ scores can be downloaded from the TBT system by the instructor in a spreadsheet format and easily uploaded into a course management system such as Blackboard. This alleviates a tremendous amount of time and money lost to photocopying, collating, stapling, distributing, collecting, and grading of paper tests.

**INSTRUCTOR USE OF TBT**

TBT is an open-source software system and can be obtained through the authors. It is very user-friendly for instructors. Students must register and choose a log in and password. Instructors then create a class and assign students to the class. Instructors also assign students to teams. As mentioned earlier, we use a stratified random sampling approach to team formation within TBL, with our considerations being a roughly equal number of males and/or student athletes on each team, along with majors. Teams consist of six or seven students. This number is the maximum that still allows for individual input and yet is the minimum to ensure an adequate number of students on test days if two or three are absent.

To create a test, the instructor simply names the test and begins adding questions. Questions and responses can be typed in or copied and pasted from a word-processing document. The instructor may choose the number of options and the correct answer. Once the test is completed, the instructor can choose from a number of options in the test manager, such as how many points each question is worth, whether items will be presented in a random or original order, and one-at-a-time or all at once with scrolling. Time limits can be set and also a warning can appear indicating a number of minutes are left. Finally, the team test can be set for how many attempts teams get to answer the question and what percentage is lost after each attempt. We usually choose five attempts (the number of options) with a 50% reduction in points after each miss. This allows students to learn the correct answer each time.

Each test, both individual and team, has its own password that is typically given to the students only when the test begins. Students are forbidden to
open any other programs during the test, such as messaging, e-mail, or web browsers to ensure item security.

Why would a student who scores very well on the individual test be motivated to do well on the team test? As with any collaborative learning task, both individual and group accountability are needed to ensure maximum effort. Test scores are a function of both individual and team performance. For teams that receive the highest score among all teams, each individual receives three bonus points. Second place teams receive two bonus points and so on. Thus, for any given test, typically fewer than half of the students receive anything for their team effort. However, we have found students to be very enthusiastic about their performance on the team test, whether this is due to the competition involved, motivation for bonus points, or simply the desire to learn the correct answers and their individual scores.

Alternately, why would a low-scoring student be motivated to participate in the team test deliberations? To prevent the “free-loader” effect, we have students provide feedback to their teammates twice during the semester. At the beginning of the semester, we introduce the teammate feedback process by saying something like:

> At the end of every semester, I usually get students outside my door who tell me they are within a few points of the next grade up, and asking if there is anything they can do for extra credit. There is not, but what I will do is look at your teammate feedback. If your feedback is good all semester long or starts out rocky and gets better, then I will consider adding a few points to bump you up. But if your teammates are requesting all semester long that you show up prepared or that you let others talk, then no, I will not consider giving you that bump.

The beauty of this, of course, is that students do not know until the very end of the semester whether they will wind up on the cusp of a given grade—so it remains a motivating force preventing everyone from free-loading, all semester long.

**LIMITATIONS OF TBT**

Although TBT affords many advantages over traditional paper-and-pencil and even IF-AT® testing, there are limitations that must be addressed. First, as with any computer-based testing, there remains the possibility that students may save screen shots of questions and copy them either to a file and export them during testing. With careful test proctoring, this possibility is slim. Second, although we have warned students about opening up other programs, we constantly see students e-mailing and web browsing in the period between the individual and team tests. In this day of cell-phone text
messaging and instant messaging, college students are very tempted to con-
tinue this practice given any free time. Again, careful test proctoring should
minimize this possibility.

Finally, as with any team or small group activity, it is possible that students
may not always work as a “team” with equal input from each team member
and equal learning that occurs as a result of the process. In a recent study
examining successful versus unsuccessful teams, Thomas (2007) found that
successful teams typically had a “superstar”—a student whose individual test
score was considerably higher than the next highest scoring team member.
This student became the “go-to-guy,” so to speak, and other team members
simply sat back and let the superstar perform. Random stratified assign-
ment in team formation is unlikely to address this problem and we are cur-
rently exploring assigning “coaches” (other students who have done well)
to low-performing teams to get them up to speed. These limitations aside,
we feel that TBT has great potential as a learning and assessment activity.

**FUTURE DIRECTIONS**

Where do we go from here? First, and most obvious, we need experimen-
tal studies that examine different aspects of TBT to determine which ones
contribute to student learning. Second, we are constantly tinkering with
TBT to further develop its potential. The authors are currently developing
software that students could use in large-scale, laptop-enabled electronic
classrooms similar to those being deployed by the Student-Centered Activi-
ties for Large Enrollment Undergraduate Programs (SCALE-UP) Project at
North Carolina State University (Beichner & Saul, 2003). The SCALE-UP
configuration (students with laptops sitting at a circular table), combined
with the TBT activity, could be used for smaller, more formative assessments
with large classes. After 10-15 min. of instruction, students could answer a
few questions and then answer the same questions with their teams to assess
understanding. This formative use of TBT would expand upon what many
teachers are currently doing with classroom performance systems (also
known as “clickers,” c.f. Lopez-Herrejon & Schulman, 2004), but would
prevent having to introduce yet another piece of technology—the clicker—
into a class that was already using TBT.

We are also exploring the addition of “word bank” functionality, so in-
structors can move beyond the multiple-choice question format. Word
banks are commonly used in language learning classes, and typically consist
of one or more pages of words or phrases that test-takers may choose from
in order to answer short-answer test questions. The experience of having to
choose a phrase or even a whole sentence from among many (perhaps doz-
ens or more) to correctly answer a question is a significantly more complex
activity than simply choosing from among four or five very short multiple-choice options. Though much more complex for the test-taker, questions answered using a word bank can nonetheless be graded “right or wrong” by the computer and scored automatically.

In closing, we are excited about the possibilities of TBT for both secondary and postsecondary classes, large and small. Laptops will become more ubiquitous and computer-based TBT will be a reasonable option even for non-tech-savvy instructors. More importantly, instructors will be able to use testing as more of a learning opportunity rather than simply an assessment opportunity.

REFERENCES


Beichner, R. J., & Saul, J. M. (2003). *Introduction to the scale-up (student-centered activities for large enrollment undergraduate programs) project*. Paper presented at the International School of Physics, Verona, Italy.


**NOTE**

1. We originally designed TBT to require that every team member must agree before submitting the team’s selection but due to technical difficulties, we decided to only have a chosen team leader submit.