

RESEARCH

Comparison of Answer-Until-Correct and Full-Credit Assessments in a Team-based Learning Course

Michelle Z. Farland, PharmD,^a Patrick B. Barlow, PhD,^b T. Levi Lancaster, PharmD,^c
Andrea S. Franks, PharmD^c

^a University of Florida College of Pharmacy, Gainesville, Florida

^b The University of Iowa Carver College of Medicine, Iowa City, Iowa

^c University of Tennessee Health Science Center College of Pharmacy, Knoxville, Tennessee

Submitted May 06, 2014; accepted August 03, 2014; published March 25, 2015.

Objective. To assess the impact of awarding partial credit to team assessments on team performance and on quality of team interactions using an answer-until-correct method compared to traditional methods of grading (multiple-choice, full-credit).

Methods. Subjects were students from 3 different offerings of an ambulatory care elective course, taught using team-based learning. The control group (full-credit) consisted of those enrolled in the course when traditional methods of assessment were used (2 course offerings). The intervention group consisted of those enrolled in the course when answer-until-correct method was used for team assessments (1 course offering). Study outcomes included student performance on individual and team readiness assurance tests (iRATs and tRATs), individual and team final examinations, and student assessment of quality of team interactions using the Team Performance Scale.

Results. Eighty-four students enrolled in the courses were included in the analysis (full-credit, n=54; answer-until-correct, n=30). Students who used traditional methods of assessment performed better on iRATs (full-credit mean 88.7 (5.9), answer-until-correct mean 82.8 (10.7), $p<0.001$). Students who used answer-until-correct method of assessment performed better on the team final examination (full-credit mean 45.8 (1.5), answer-until-correct 47.8 (1.4), $p<0.001$). There was no significant difference in performance on tRATs and the individual final examination. Students who used the answer-until-correct method had higher quality of team interaction ratings (full-credit 97.1 (9.1), answer-until-correct 103.0 (7.8), $p=0.004$).

Conclusion. Answer-until-correct assessment method compared to traditional, full-credit methods resulted in significantly lower scores for iRATs, similar scores on tRATs and individual final examinations, improved scores on team final examinations, and improved perceptions of the quality of team interactions.

Keywords: team-based learning, active learning, pharmacy education, answer-until-correct, partial credit

INTRODUCTION

A variety of methods are used to grade assessments in courses that use team-based learning (TBL). Pharmacy educators use of the Immediate Feedback Assessment Technique, or the IF-AT (Epstein Educational Enterprises, Cincinnati, OH), on tRATs and on team examinations.¹⁻⁷ In Allen and colleagues survey of faculty members from colleges' and schools of pharmacy in the United States, 74% of respondents who had implemented TBL were using the IF-AT forms for assessment.⁸ The IF-AT card

is a response form with a preset answer key designed to be used with multiple-choice questions.⁹⁻¹¹ Students work in teams to determine the best answer to an assessment item and scratch the area under the answer they think is correct. A star is revealed when students answer correctly, and a blank box is revealed when students answer incorrectly. Following an incorrect answer, students have the opportunity to scratch the area under another answer choice until the correct answer is revealed. This method of answer-until-correct allows for partial credit to be awarded for each item of the assessment. It also provides immediate feedback for students and faculty members.

Of those who report using IF-AT forms in TBL sessions, only 2 specify using a partial grading method.³⁻⁴

Corresponding Author: Michelle Z. Farland, PharmD, 1225 Center Drive, HPNP 3307, PO Box 100486, Gainesville, FL 32610-0486. Tel: 352-273-6293. Fax: 352-273-6242. E-mail: mfarland@cop.ufl.edu

However, the grading method is not compared to other methods in these reports. Persky and colleagues assessed the impact of IF-AT forms compared to traditional assessment methods (ie, multiple-choice, true/false, short-answer) for individual course examinations in a pharmacokinetics course not using TBL.¹² They found no differences in student performance but did find that students preferred IF-AT forms over traditional assessment methods. Published comparisons of student performance in pharmacy education using IF-AT forms for team assessments in TBL courses do not describe the impact on the quality of team interactions.

The purpose of this study was to assess the impact of awarding partial credit to team assessments using an answer-until-correct method compared to traditional methods of grading on both student performance and quality of team interactions.

METHODS

Team-based learning was implemented in 2008 in an elective course titled Ambulatory Care.¹³ Since this time, the course has been offered annually during the fall semester to third-year student pharmacists and has used TBL as the sole instructional method. The course directors consistently incorporate the core elements of TBL in the course structure.¹⁴ During the 2013 fall semester, IF-AT forms were introduced to the course to be used as the assessment tool for tRATs.

We conducted a retrospective longitudinal study that included students enrolled in the ambulatory care elective in the 2010, 2011, and 2013 fall semesters. The control group (full-credit) consisted of students enrolled in the course in the 2010 and 2011 fall semesters. The course used a traditional, full-credit assessment method for tRATs at those times. Students from those course offerings were identified as the control group because they had completed the Team Performance Scale as a component of the course to assess quality of team interactions. The intervention group (answer-until-correct) consisted of students enrolled in the course in the 2013 fall semester when IF-AT forms were used to grade tRATs. Topics included in each course offering varied slightly from year to year based on instructor availability. Table 1 displays the topics included in each course offering. Topics were consistent in the course except for anticoagulation part 2, women's health, and chronic noncancer pain. Students were divided into 5-6 member teams. At the start of each class session, students completed an iRAT, and immediately after, the tRAT. The topic instructor then led a discussion of misconceptions identified by the tRAT. Students completed the team application activity, which was followed by interteam discussions led by the topic

instructor. Students completed one course examination at the conclusion of the course, first as individuals, then as teams. All assessments comparing individual and team performance used the same items.

All iRATs and tRATs consisted of 10 multiple-choice questions with 4 answer choices. The total possible points earned on tRATs changed when the answer-until-correct method was implemented. The full-credit group could earn up to 10 possible points on the tRAT, 1 point per question. The answer-until-correct group could earn up to 40 possible points on the tRAT. Partial credit was awarded based on the number of answer choices revealed on the IF-AT form. Full credit (4 points) was awarded when students revealed 1 answer choice. Half credit (2 points) was awarded when students revealed 2 answer choices. Students who revealed 3 and 4 answer choices were awarded 1 and 0 points, respectively. Grades from the answer-until-correct group's tRATs were scaled to a 10-point scale for comparison.

Study outcomes included student performance on iRATs, tRATs, individual final examination, team final examination, and student-rated quality of team interactions using the Team Performance Scale, an 18-item survey that has evidence for reliability (Cronbach $\alpha=0.97$) to assess the quality of team interactions.¹⁵ Each item was assessed on a 0-6 Likert scale with 0 representing none of the time and 6 representing all of the time (highest possible composite score was 108). Higher composite scores indicated higher performing teams.

Descriptive statistics were used to summarize student performance on the iRAT and tRAT by topic and overall and on the Team Performance Scale. For statistical comparison, number of correct items (raw score) was used as opposed to percent correct. Total possible points

Table 1. Topics Addressed in the Course by Semester

Topic	Fall 2010	Fall 2011	Fall 2013
Diabetes Part 1	x	x	x
Diabetes Part 2	x	x	x
Hypertension	x	x	x
Heart Failure	x	x	x
Dyslipidemia	x	x	x
Anticoagulation Part 1	x	x	x
Anticoagulation Part 2	x		x
Asthma	x	x	x
Chronic Obstructive Pulmonary Disease	x	x	x
Tobacco Cessation	x	x	x
Chronic Noncancer Pain	x	x	
Women's Health		x	x
Total Number of Topics	11	11	11

for the iRAT and tRAT were 110 points each (1 point per item, 10-item RATs, 11 RATs in each course). Total possible points earned for the individual and team final examinations were 50 points each (1 point per item, 50-item examination). Preliminary Pearson correlations were conducted to assess whether cumulative grade point average (GPA) was significantly associated to any of the outcome variables, and would therefore need to be controlled as a possible confounder. When such a confounding relationship was found, analysis of covariance (ANCOVA) was performed to compare the 2 study groups' performance on each outcome while controlling for cumulative GPA. Independent *t* tests were used to compare the study groups when no confounding effect was discovered. All tests were 2-sided, and an alpha value below 0.05 was considered significant. Prior work established reliability of the Team Performance Scale in a sample of medical students.¹⁵ Cronbach alpha was calculated as a measure of reliability for the Team Performance Scale in a sample of student pharmacists to confirm it could be used in this population. Analyses were conducted using IBM SPSS v.21 (IBM Inc., Chicago, IL). The study was deemed exempt by the Institutional Review Board.

RESULTS

Eighty-four students enrolled in the courses were included in the analysis (full credit, n=54; answer-until-correct, n=30). Cumulative GPA prior to enrollment in the course was similar between groups (full-credit mean 3.1 (0.3), answer-until-correct mean 3.1 (0.4), *p*=0.693).

However, GPA was significantly related to both iRAT and tRAT scores, so it was controlled for in these analyses. The average iRAT scores are listed in Table 2. Students enrolled in the full-credit course offering scored significantly higher on iRATs overall (full credit mean 88.7 (5.9), answer-until-correct mean 82.8 (10.7), *p*<0.001) with differences found on the following topics: heart failure, hypertension, anticoagulation part 2, women's health, smoking cessation, and dyslipidemia. The average tRAT scores are listed in Table 3. There was no difference between study groups in overall performance on the tRAT (full-credit mean 105.0 (2.9), answer-until-correct mean 104.7 (3.9), *p*=0.635). However, there was a significant difference found in student performance on tRATs for heart failure, hypertension, anticoagulation part 2, asthma, women's health, and dyslipidemia.

There was no statistical difference found in student performance on the individual final examination (full-credit mean 38.0 (3.7), answer-until-correct mean 39.4 (4.4), *p*=0.147). Students in the answer-until-correct group scored significantly higher on the team final examination (full-credit mean 45.8 (1.5), answer-until-correct 47.8 (1.4), *p*<0.001).

The reliability testing using Cronbach alpha indicated the team performance scale was a sufficiently reliable instrument (Cronbach α =0.94). Quality of team interactions differed significantly between study groups (full-credit 97.1 (9.1), answer-until-correct 103.0 (7.8), *p*=0.004). Table 4 summarizes the results for each item of the Team Performance Scale. Statistical comparison of each item was not conducted as the tool was not intended

Table 2. Individual Readiness Assessment Test Scores by Topic and Study Group

Topic ^a	Full-Credit		Answer-Until-Correct		<i>p</i> value ^b
	<i>n</i>	Mean (SD)	<i>n</i>	Mean (SD)	
Heart Failure	54	8.8 (1.2)	30	7.9 (1.7)	0.014
Hypertension	54	8.4 (1.3)	30	7.4 (1.5)	0.001
Diabetes Part 1	54	7.5 (1.7)	30	6.9 (2.2)	0.088
Diabetes Part 2	54	6.9 (1.4)	30	7.1 (1.4)	0.604
Anticoagulation Part 1	54	7.9 (1.3)	30	7.5 (1.6)	0.262
Anticoagulation Part 2	28	9.6 (0.5)	30	6 (2.6)	<0.001
Asthma	54	8.7 (1.2)	29	8.6 (1.7)	0.769
Chronic Obstructive Pulmonary Disease	54	8.7 (1.0)	30	8.4 (1.4)	0.234
Women's Health	26	6.7 (1.7)	30	8.9 (0.9)	<0.001
Tobacco Cessation	54	7.7 (1.4)	30	7.1 (1.8)	0.036
Dyslipidemia	54	8.7 (1.2)	30	7.5 (1.5)	<0.001
Overall	54	88.7 (5.9)	30	82.8 (10.7)	<0.001

^a Chronic noncancer pain was not included as a separate item as it was not offered in the course that used answer-until-correct method. However, scores from this topic were included in the calculation of the overall comparison.

^b Comparisons assessed using analysis of covariance controlling for cumulative GPA (hypertension, diabetes part 1 and 2, anticoagulation part 2, smoking cessation, women's health, dyslipidemia, overall), or independent *t* tests (heart failure, anticoagulation part 1, asthma, chronic obstructive pulmonary disease).

Table 3. Team Readiness Assessment Test Scores by Topic and Study Group

Topic ^a	Full Credit		Answer-Until-Correct		p value ^b
	n	Mean (SD)	n	Mean (SD)	
Heart Failure	54	9.9 (0.3)	30	9.7 (0.4)	0.002
Hypertension	54	9.8 (0.4)	30	9.5 (0.5)	0.004
Diabetes Part 1	54	9.4 (1.0)	30	9.8 (0.3)	0.055
Diabetes Part 2	54	9.2 (0.7)	30	9.4 (0.5)	0.128
Anticoagulation Part 1	54	9.4 (0.8)	30	9.5 (0.4)	0.344
Anticoagulation Part 2	28	10 (0)	30	9.2 (1.0)	<0.001
Asthma	54	9.6 (0.7)	30	9.9 (0.2)	0.004
Chronic Obstructive Pulmonary Disease	54	9.9 (0.3)	30	9.8 (1.2)	0.475
Women's Health	26	9.2 (0.8)	30	10 (0)	<0.001
Tobacco Cessation	54	9.3 (0.8)	30	9.3 (0.5)	0.921
Dyslipidemia	54	10 (0)	30	9.2 (0.9)	<0.001
Overall	54	105.0 (2.9)	30	104.7 (3.9)	0.635

^a Chronic noncancer pain was not included as a separate item as it was not offered in the course that used answer-until-correct method. However, scores from this topic were included in the calculation of the overall comparison.

^b Comparisons assessed using analysis of covariance controlling for cumulative GPA (diabetes part 1, anticoagulation part 2, smoking cessation, and overall), or independent t-tests (heart failure, hypertension, diabetes part 2, anticoagulation part 1, asthma, chronic obstructive pulmonary disease, women's health, and dyslipidemia).

to be used in this manner. However, the mean score reported for each item was higher in the answer-until-correct group.

DISCUSSION

Student performance was mixed depending on type of assessment and method of grading. Students performed significantly better on iRATs when using traditional methods of grading for tRATs. This contradicted the hypothesis that different methods of grading team assessments would not impact performance on individual assessments. One theory why this may have occurred is that student preparation prior to class sessions may have declined when partial credit was available for team assessments. Part of the rationale behind using both an individual and team assessment prior to the application phase of TBL is that students are held accountable to themselves in the individual assessment and to the team in the team assessment. This process may entice students to prepare more thoroughly before class to not let themselves or the team down. Thus, when partial credit on team assessments is introduced, some pressure for team accountability is potentially alleviated, resulting in less optimal preclass preparation; however, this theory needs to be scientifically tested.

Overall student performance on tRATs was similar between grading methods used. However, students in the full-credit group performed significantly better on 4 of the 11 course topics, whereas the partial-credit group performed better on 2 of the 11 course topics. This outcome was unexpected, as one would assume the opportunity for partial credit would have potential to increase the grade

for the assessment, and indicates that use of partial credit using an answer-until-correct method did not inappropriately inflate grades for the tRAT. There were, however, anecdotal observations made by course faculty members that students who used the answer-until-correct method spent more time in team discussion while completing the tRAT compared to students in prior course offerings. Persky and colleagues also observed that students required additional time to complete assessments when using the answer-until-correct method for individual course examinations.¹² Prior to the implementation of IF-AT forms in the pharmacokinetics course, examinations consisted of open-ended problems, multiple-choice, true/false, and short-answer questions and were scheduled for 2 hours. Following implementation of the IF-AT forms, the entire examination was made up of multiple-choice questions, and its duration increased to 3 hours to allow sufficient time for completion.

Student performance on individual course examinations did not differ between groups as was expected. Performance on the team course examination was significantly higher in the answer-until-correct group. One possible reason for the observed but small difference in team final examination performance was the emphasis on team development throughout the course that led to improved team performance. The Team Performance Scale scores were higher for teams that used the answer-until-correct assessment method compared to the full-credit group. However, overall tRAT scores did not differ significantly between methods. Future research should be conducted to determine if either assessment method used in this study

Table 4. Team Performance Scale Item Summary

Team Performance Scale Item ($\alpha=0.94$)	Mean (SD) ^a	
	Full Credit	Answer-Until-Correct
1. All team members made an effort to participate in discussions.	5.4 (0.7)	5.7 (0.7)
2. When team members had different opinions, each member explained his or her point of view.	5.4 (0.8)	5.8 (0.5)
3. Team members encouraged one another to express their opinions and thoughts.	5.4 (0.8)	5.9 (0.4)
4. Team members shared and received criticism without making it personal.	5.4 (0.8)	5.7 (0.7)
5. Different points of view were respected by team members.	5.4 (0.7)	5.9 (0.3)
6. Often members helped a fellow team member to be understood by paraphrasing what he or she was saying.	5.2 (0.9)	5.6 (0.9)
7. My team used several techniques for problem solving (such as brainstorming) with each team member presenting his or her best ideas.	5.2 (0.9)	5.5 (0.9)
8. Team members worked to come up with solutions that satisfied all members.	5.4 (0.7)	5.7 (0.7)
9. All team members consistently paid attention during group discussions.	5.0 (0.9)	5.5 (0.8)
10. My team actively elicited multiple points of view before deciding on a final answer.	5.4 (0.8)	5.8 (0.6)
11. Team members listened to each other when someone expressed a concern about individual or team performance.	5.5 (0.6)	5.8 (0.4)
12. Team members willingly participated in all relevant aspects of the team.	5.4 (0.8)	5.8 (0.7)
13. Team members resolved differences of opinion by openly speaking their mind.	5.5 (0.6)	5.8 (0.6)
14. Team members used feedback about individual or team performance to help the team be more effective.	5.3 (0.8)	5.8 (0.6)
15. Team members seemed attentive to what other team members were saying when they spoke.	5.4 (0.6)	5.8 (0.6)
16. My team resolved many conflicts by compromising between team members, with each one giving in a little.	5.5 (0.6)	5.7 (0.8)
17. Members who had different opinions explained their point of view to the team.	5.5 (0.6)	5.9 (0.4)
18. Team members were recognized when something they said helped the team reach a good decision.	5.5 (0.7)	5.8 (0.6)

^a Each item rated on 6-point scale from 0=none of the time to 6=all of the time.

could help teams work through Tuckman's stages of team development (forming, storming, norming, performing).¹⁶ It can be hypothesized that teams who reach the performing stage would perform better on team activities such as a tRAT or team final examination.

A potential limitation of the study was that topics and faculty facilitators changed slightly in the course each year, although 80% of the topics and faculty facilitators remained the same. Also, differences in RAT items among courses might have contributed to the differences observed in iRAT and final team examination scores. While the content was similar in each course offering, the RAT questions did not remain exactly the same, based on feedback from students during the appeal process that faculty members used to improve questions for future offerings. This may have limited the ability to compare findings from different offerings of the same course. However, constant revisions and updates to course content reflect the reality of teaching practice. Another limitation was that the sample size was relatively small, but the analysis was conducted in an elective course with a smaller enrollment

than required courses in the curriculum. A similar comparison in a course with higher student enrollment should be conducted to confirm the results of this study.

CONCLUSION

The answer-until-correct assessment method compared to traditional methods using full-credit resulted in lower iRAT scores, similar tRAT and individual final examination scores, and improved team final examination scores and perceptions on team interaction quality. Additional research is needed to identify reasons for decreased performance on iRATs and to assess the impact answer-until-correct assessment methods have on time to complete the assessment. We also recommend that additional research focus on assessing progression through stages of team development to help identify activities that assist teams in reaching the performing stage.

REFERENCES

1. Persky AM. The impact of team-based learning on a foundational pharmacokinetics course. *Am J Pharm Educ.* 2012;72:Article 31.

American Journal of Pharmaceutical Education 2015; 79 (2) Article 21.

2. Persky AM, Pollack GM. A modified team-based learning physiology course. *Am J Pharm Educ.* 2011;75:Article 204.
3. Kolluru S, Roesch DM, de la Fuente AA. A multi-instructor, team-based, active-learning exercise to integrate basic and clinical sciences content. *Am J Pharm Educ.* 2012;76:Article 33.
4. Redwanski J. Incorporating team-based learning in a drug information course covering tertiary literature. *Curr Pharm Teach Learn.* 2012;4(3):202-206.
5. Pogge E. A team-based learning course on nutrition and life-style modification. *Am J Pharm Educ.* 2013;77(5):Article 103.
6. Grady SE. Team-based learning in pharmacotherapeutics. *Am J Pharm Educ.* 2011;75(7):Article 136.
7. Bleske BE, Remington T, Wells TD, et al. Team-based learning to improve learning outcomes in a therapeutics course sequence. *Am J Pharm Educ.* 2014;78(1):Article 13.
8. Allen RE, Copeland J, Franks AS, et al. Team-based learning in US colleges and schools of pharmacy. *Am J Pharm Educ.* 2013;77(6): Article 115.
9. Epstein ML, Epstein BB, Brosvic GM. Immediate feedback during academic testing. *Psychol Rep.* 2001;88(3 Pt1):889-894.
10. Epstein ML, Brosvic GM. Students prefer the immediate feedback assessment technique. *Psychol Rep.* 2002;90(3 Pt2):1136-1138.
11. Parmelee D, Michaelsen LK, Cook S, Hudes P. Team-based learning: a practical guide: AMEE Guide No. 65. *Med Teach.* 2012;34:e275-e287.
12. Persky AM, Pollack GM. Using answer-until-correct examinations to provide immediate feedback to students in a pharmacokinetics course. *Am J Pharm Educ.* 2008;72(4):Article 83.
13. Zingone MM, Franks AS, Guirguis AB, George CM, Howard-Thompson A, Heidel RE. Comparing team-based and mixed active-learning methods in an ambulatory care elective course. *Am J Pharm Educ.* 2010;74(9):Article 160.
14. Farland MZ, Sicat BL, Franks AS, Pater KS, Medina MS, Perksy AM. Best practices for implementing team-based learning in pharmacy education. *Am J Pharm Educ.* 2013;77(8):Article 177.
15. Thompson BM, Levine RE, Kennedy F, et al. Evaluating the quality of learning-team processes in medical education: development and validation of a new measure. *Acad Med.* 2009;84(10 Suppl):S124-S127.
16. Tuckman BW. Developmental sequence in small groups. *Psychol Bull.* 1965;63(6):384-399.