RURALS: Review of Undergraduate Research in Agricultural and Life Sciences

Volume 6 | Issue 1 Article 1

March 2012

Favorable Team Scores Under the Team-Based Learning Paradigm: A Statistical Artifact?

Trevor Hefley
University of Nebraska Lincoln, trevorhefley@msn.com

Andrew J. Tyre
University of Nebraska at Lincoln, atyre2@unl.edu

Favorable Team Scores Under the Team-Based Learning Paradigm: A Statistical Artifact?

Cover Page Footnote

We thank Erin E. Blankenship for her kind assistance with statistical matters. We also thank Mark E. Burbach, Gina S. Matkin, and Jamie McFadden for their constructive and helpful reviews. Trevor J. Hefley graduated in 2010 with a bachelors of science in fisheries and wildlife at the University of Nebraska-Lincoln. He is now a PhD student in the Department of Statistics and School of Natural Resources at the University of Nebraska-Lincoln. Dr. Andrew J. Tyre is an associate professor of wildlife ecology in the School of Natural Resources at the University of Nebraska-Lincoln. Review was coordinated by Steve Kachman, professor and interim chair, Department of Statistics, University of Nebraska-Lincoln.

1. Introduction

Many higher educators seek non-traditional methods for instruction to enhance student learning experience. One of the most commonly employed non-traditional methods of instruction is active small group learning which includes team-based learning (TBL), cooperative learning, and problem-based learning (Jones & Jones 2008; Michaelsen, Knight, & Fink, 2002; Nicoll-Senft, 2009). Of these three instructional methods, TBL is the most highly structured and promotes mastering of learning objectives using readiness assurance testing (RAT). Readiness assurance testing is a sequence of individual and team based tests that are administered regularly and cover material that is to be learned outside of class. At the beginning of an instructional period, individual readiness assurance tests (IRAT) are administered to individual students. Typically the IRAT consists of a short multiple-choice test, but can follow other schemes such as short answer or fill in the blank. After the IRAT, individuals gather into permanently assigned teams to take a team readiness assurance test (TRAT). The TRAT consists of the same questions as the IRAT, but is followed by immediate feedback that is provided after each attempt by the group to answer a question. Immediate feedback provides the opportunity for groups to discuss and prepare an alternative answer. Such small group discussion is central to the TBL paradigm. To promote discussion instructors often use corrective scoring by providing partial credit for subsequent attempts at incorrectly answered questions on the TRAT (Michaelsen et al., 2002; Michaelsen, Parmelee, McMahon, & Levine, 2008).

There are many techniques to administer RATs (Gomez, Dehzhi, & Katia, 2010; Michaelsen et al., 2002; Robinson & Walker, 2008). A technique promoted by Michaelsen et al. (2002) and used in other case studies of TBL is a multiple choice format for the IRAT and an immediate feedback assessment technique (IF-AT) for the TRAT (Carmichael, 2009; Clark, 2008; Cotner, Baepler, & Kellerman, 2009; Haberyan, 2007). The IF-AT format is similar to multiple choice formats, but allows for corrective feedback in that the form is similar to a "scratch off" lottery ticket and groups are able to make multiple attempts until the correct answer is uncovered (Epstein, Epstein, & Brosvic, 2001). Using corrective feedback contributes to longer retention, promotes group discussion, and allows for instructors to award partial credit for proximal knowledge (Epstein et al., 2001; Michaelsen et al., 2002; Michaelsen et al., 2008).

A typical multiple choice IRAT does not allow for corrective feedback. Michaelsen et al. (2002) suggests using a point spreading system, when IF-AT forms are not in use (i.e. the IRAT). Under point spreading, each multiple choice

question is worth 3 points and the students can spread the 3 points across multiple answers if uncertain of the correct answer or allocate all 3 points to a single answer if they are certain. If instructors or students want to compare IRAT and TRAT scores to evaluate the contribution of team learning, it is imperative that statistically equivalent scoring schemes are used when using two different techniques for the IRAT and TRAT.

Su (2004) found that the greater the positive difference between TRAT scores and IRAT scores, the more the students preferred working in teams. If different scoring techniques are used between the IRAT and TRAT it is ambiguous for students to analyze the contribution of their team and may lead to either a negative or overly positive view of team contribution (Su, 2004). In contrast, instructors may conclude that TRAT scores are higher or lower than IRAT scores when in reality the two scores can not be compared because they are not statistically equivalent (Michaelsen et al., 2002; Michaelsen et al., 2008). Furthermore, a statistically equivalent scoring technique will ensure students perceive grading to be fair and equitable, a guiding principle of TBL (Michaelsen et al., 2002; Michaelsen & Sweet, 2008).

We conducted this study to determine the pervasiveness of multiple choice and IF-AT scoring inconsistencies in the TBL literature, highlight the statistical difference between common methods, and offer suggestions for corrective scoring techniques.

2. Method

Literature review

We reviewed published literature sources, including books that described the methods used by the instructor to administer TBL. We used Academic Search Premier and searched using the following key words either individually or in various combinations: team-based learning, IRAT, TRAT and IF-AT. We compiled our results into two categories: studies that used multiple choice testing and IF-AT forms and studies that did not.

Analysis

An expected value is the statistical mean of a probability distribution. The expected value for discrete distributions is calculated by summing, across all questions, the product of the probability a student gets the question correct by the point value of each question (Wackerly, Mendenhall, & Scheaffer, 2008). The expected values can be converted into a percent equivalent by dividing the ex-

pected value by the total number of points possible for each question. We calculated the percent equivalent expected value of 4-choice single answer multiple choice IRAT

$$E_{IRAT}(P) = 1P + 0(1-P) \quad (1)$$

$$E_{IRAT}(P) = P, \quad (2)$$

were *P* is the probability of selecting the correct answer on the first attempt. We also calculated the percent equivalent expected value of 4-choice IF-AT TRAT where the point value of the question is decremented by 25% for each attempt to answer the question after the first (Michaelsen et al., 2002)

$$E_{TRAT}(P) = P + \frac{3}{4}\left(\frac{1-P}{3}\right) + \frac{1}{2}\left(\frac{1-P}{3}\right) + \frac{1}{4}\left(\frac{1-P}{3}\right) \quad (3)$$

$$E_{TRAT}(P) = \frac{P+1}{2}. \quad (4)$$

Both expected values and percent equivalents of the IRAT and TRAT are a function of the probability of selecting the correct answer on the first attempt. Since the expected value is a function of *P*, by our definition, *P* is a measure of individual and team ability. We assumed that if the student(s) were incorrect on their first attempt that they selected from the remaining possible answers at random for both IRAT's and TRAT's. This is analogous to students guessing on subsequent answers if the students are incorrect on the first attempt. We calculated the percent equivalent difference in expected values as a function of *P* for the two techniques

$$E_{TRAT}(P) - E_{IRAT}(P) = \frac{P_{TRAT}+1}{2} - P_{IRAT} \quad (5)$$

$$E_{TRAT}(P) - E_{IRAT}(P) = \frac{1-P}{2} \quad If\ P_{TRAT} = P_{IRAT}, \quad (6)$$

were P_{TRAT} and P_{IRAT} is the probability of selecting the correct answer on the first attempt for teams and individual respectively. We then calculated the percent equivalent expected difference under the three scenarios: where *P* is the same, 10% greater, and 10% less for TRAT's than IRAT's respectively. Given the various schemes of multiple choices testing, such as point spreading, we limited our

analysis to simple single answer multiple choice. It should be noted that the expected value of single answer multiple choice maximizes the expected value when compared to multiple choice schemes that allow for point spreading. Single answer multiple choice is equivalent to point spreading in the case where a student puts all points on a single answer. In sum our results minimize the difference between TRAT and IRAT scores; alternative assumptions lead to greater discrepancies between TRAT and IRAT expected scores.

3. Results and Discussion

We identified 14 published studies that provided descriptions of TBL scoring schemes. Of the 14 identified studies, 21% used multiple choice and IF-AT scoring techniques. We also reviewed current books which contained case studies authored by instructors. Of 18 case studies, 33% used multiple choices and IF-AT scoring techniques, 44% used other techniques, and 22% used unknown. Overall the use of multiple choice and IF-AT scoring techniques was 28%.

Table 1. Categorization of published Team Based Learning case studies based on two scoring criteria: studies that used multiple choice and IF-AT scoring techniques and studies that used other scoring methods.

Citation	Multiple choice and IF-AT scoring techniques	Other
Carmichael (2009) ..	x	
Chung et al. (2009) .		x
Clark et al. (2008)	x	
Dunaway (2005)		x
Haberyan (2007)	x	
Koles et al. (2005)		x
Levine et al. (2004)		x
MacPherson & Bruecker (2008)		x
Mcinerney & Fink (2003)		x
Nicoll-Senft (2009)		x
Nieder et al. (2004)		x
Shellenberger (2009)		x
Touchet & Coon (2005)		x
Weiner et al. (2009)		x

Using the scoring scheme of 4-choice single answer multiple choice IRAT and 4-choice IF-AT TRAT techniques we showed that in almost all cases the percent difference in TRAT and IRAT scores was positive, indicating that under most circumstances teams will score higher than individuals even if individual ability is greater than team ability (Figure 1). For our three scenarios the difference in expected value of TRAT and IRAT scores decreased as P increased. For the scenario where $P_{TRAT} = P_{IRAT}$, the difference is always positive, until P was equal to 1 at which point the difference was nil. For the scenario where P_{TRAT} was 10% lower than P_{IRAT}, the expected value of the difference was positive until P_{TRAT} was equal to 0.8. Likewise, for the scenario where P_{TRAT} was 10% greater than P_{IRAT} the expected value of the difference was always positive.

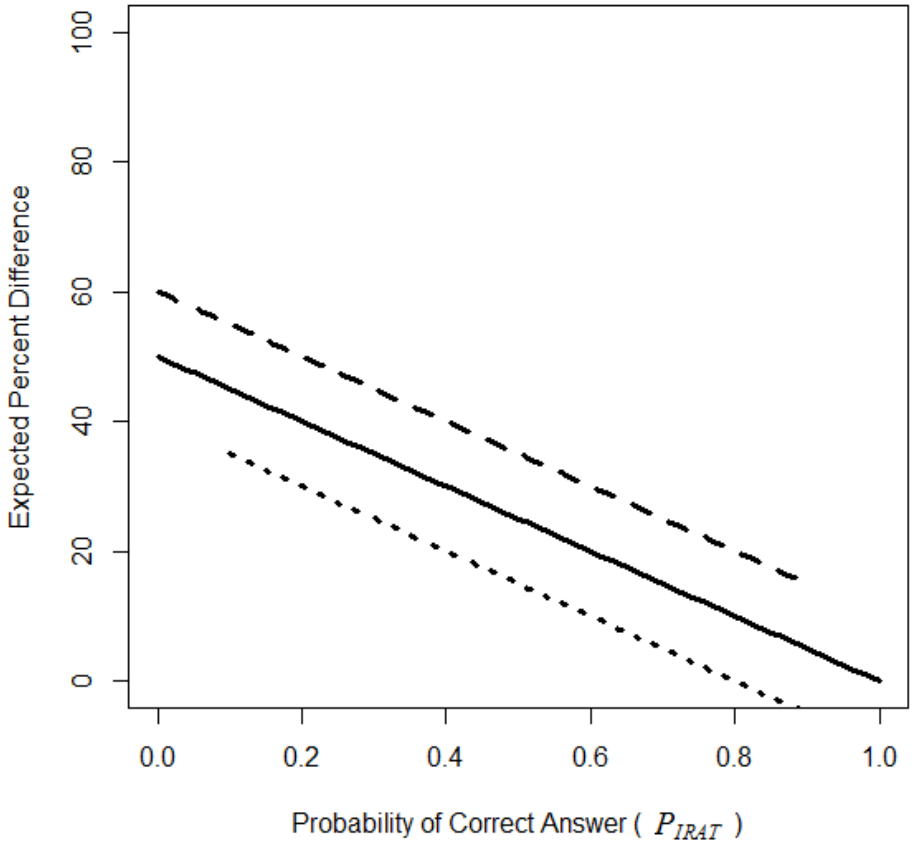

Figure 1. Expected percent difference in IF-AT and multiple choice grading techniques (TRAT minus IRAT score) as a function of the probability (P_{IRAT}) of answering the first question correctly for individual students under three scenarios: where the probability of selecting the correct answer is the same (solid middle line), 10 % greater (dashed top line), and 10% less (dotted bottom line) for TRAT than IRAT respectively.

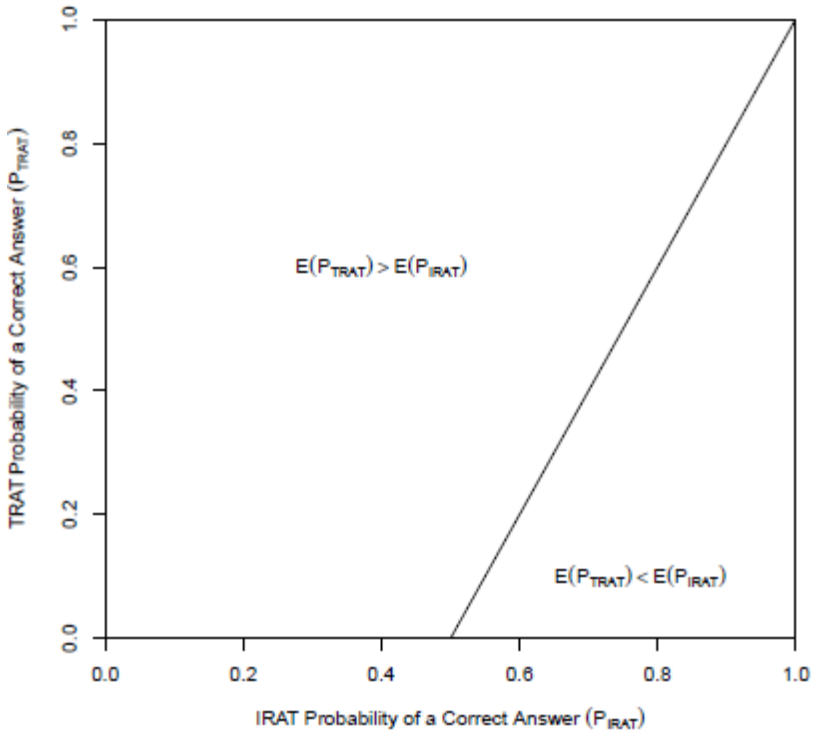

Figure 2. Line of equal expected percent grade as a function of the probability of a correct answer for the IRAT and TRAT if IF-AT and multiple choice grading techniques are used.

We found that multiple choice and IF-AT scoring techniques were common among published TBL studies. Although the studies that used multiple choice and IF-AT scoring techniques may have used different scoring schemes than we used in our analysis, the general results hold, as no author offered statistical correction or acknowledgement of the anomaly we have shown.

Our results show that with the scoring scheme of single answer multiple choice IRAT and IF-AT TRAT techniques, under almost all circumstances, TRAT scores will be greater than IRAT scores (Figure 2). For example if students were to randomly guess on the IRAT and TRAT (i.e. $P_{TRAT} = P_{IRAT} = 0.25$) the students would score 37.5% higher on the TRAT compared to the IRAT. However, the percent discrepancy we have shown in IRAT and TRAT scores depends on the probability of selecting a correct answer on the first attempt (P)and the assumption of equal chances of selecting wrong answers for both multiple choice and IF-AT techniques.

One of the tenets of TBL is that teams perform better than individuals (Michaelsen et al., 2002). Increased team performance would suggest the proba-

bility of selecting a correct answer on the first attempt would be greater for teams (i.e. $P_{TRAT} > P_{IRAT}$). When $P_{TRAT} > P_{IRAT}$, we showed the expected value of the TRAT will always be greater than the expected value of the IRAT. It is unlikely that $P_{TRAT} < P_{IRAT}$, hence under the scoring scheme of multiple choice IRAT and IF-AT TRAT, TRAT scores will always be greater than IRAT scores. In general, when $P_{TRAT} > P_{IRAT}$, TRAT scores will be greater than IRAT scores, but the magnitude of the difference will depend on P_{TRAT}, P_{IRAT}, and the assumptions of our analysis. We do not know if the assumptions of equal chances of selecting a wrong answer for both multiple choice and IF-AT techniques is correct, but our general results will hold under normal condition, that is, unless P_{TRAT} is relatively small compared to P_{IRAT} and the probability of alternative answers being correct on IF-AT forms is exceedingly high.

4. Implications

Instructors may want to evaluate the effectiveness of TBL by comparing IRAT and TRAT scores. In addition, the desire of students to work in teams is partially motivated by the perceived increase in the team's ability over the individual (Su, 2004). With multiple choice and IF-AT grading scheme used in 28% of TBL studies and recommended by leading TBL sources, teams by default will score higher. Under this scoring scheme IRAT and TRAT scores can not be compared without tedious statistical correction. We suggest that instructors use equivalent scoring schemes for IRAT and TRAT testing. If instructor uses IF-AT forms for the TRAT, the equivalent multiple choice IRAT would be a ranking of answers from best to worst. For example, if each question was worth 4 points, IRAT testing would involve numerically ranking the 4 possible answers with 4 being the best and 1 being the worst; the student would receive the point value rank for the correct answer. Similarly, the TRAT using IF-AT forms would be scored by allocating 4 points if the correct answer is obtained on the first attempt, 3 if the correct answer is obtained on the second attempt, etc. Although higher TRAT scores occur under statistically equivalent scoring schemes when $P_{TRAT} > P_{IRAT}$, the magnitude of the difference between TRAT and IRAT scores is constant for a constant difference in P. With statistically equivalent scoring the difference in TRAT and IRAT scores represents increased team performance over individual performance.

Although statistically equivalent, the methods we describe above still may not lead to directly comparable IRAT and TRAT scores as teams may be able to identify alternative answers that are correct based on immediate feedback. The proba-

bility of correctly selecting alternative answers using IF-AT techniques may be conditional on knowing the correctness of each attempt. Although difficult to estimate, it is likely that this bias would favor TRAT scores as no information (i.e. immediate feedback) is provided for the IRAT that could increase the probability of correctly selecting alternative answers.

We have applied TBL in our classroom and observed other faculty apply TBL. One of the difficulties we have encountered is the scoring inconsistency described above. In our experience students do not see the IRAT and TRAT formats as equivalent. Such scoring inconsistencies have decreased student satisfaction with TBL and caused scoring changes mid-way through the semester. Rightfully so, students lost confidence in and resisted TBL following modification of an already complex and foreign grading system. We have found such scoring inconsistencies both prevalent in the literature and in practice at our own institution. We feel that in order for TBL to be successful, accepted by students, and functional for instructors, the IRAT and TRAT scores must be easily comparable.

5. References

References marked with an asterisk indicate studies included in the literature review

*Carmichael, J. (2009). Team-based learning enhances performance in introductory biology. *Journal of College Science Teaching*, 38(4), 54–61.

*Chung, E. K., Rhee, J. A., Baik, Y. H. & A, O. S. (2009). The effect of team-based learning in medical ethics education. *Medical Teacher*, 31, 1013–1017.

*Clark, M. C., Nguyen, H. T., Bray, C., & Levine, R. E. (2008). Team-based learning in an undergraduate nursing course. *Journal of Nursing Education*, 47, 111–117.

Cotner, S., Baepler, P., & Kellerman, A. (2008). Scratch this! The IFAT as a technique for stimulating group discussion and exposing misconceptions. *Journal of College Science Teaching*, 37 (4), 48–53.

*Dunaway, G. A. (2005). Adaption of team learning to an introductory graduate pharmacology course. *Teaching and Learning in Medicine*, 17, 56–62.

Epstein, M. L., Epstein, B. B., & Brosvic, G. M. (2001). Immediate feedback during academic testing. *Psychological Reports*, 88, 889–894.

Gomez, E. A., Dezhi, W., and Katia, P. (2010). Computer-supported team-based learning: the impact of motivation, enjoyment and team contributions on learning outcomes. *Computers and Education*, 55, 378–390.

*Haberyan, A., (2007). Team-based learning in an industrial/organizational psychology course. *North American Journal of Psychology*, 9,143–152.

Jones, K. A., and J. L. Jones. (2008). Making cooperative learning work in the college classroom: an application of the 'Five Pillars' of cooperative learning to post-secondary instruction. *Journal of Effective Teaching*, 8, 61–76.

*Koles, P., Nelson, S., Stolfi, A., Parmelee D., & DeStephen D. (2005). Active learning in a year 2 pathology curriculum. *Medical Education*, 39, 1045–1055.

*Levine, R. E, Boyle, M. O., Haidet, P., Lynn D. J., Stone, M. M., Wolf, D. V., et al. (2004). Transforming a clinical clerkship with team learning. *Teaching and Learning in Medicine*, 16, 270–275.

*MacPherson, B. R., & Bruecker, J. K. (2008). Using modified team-based learning in the medical histology lab. *The FASEB Journal*, 22, 236.2.

*Mcinerney, M. J. & Fink, L. D. (2003). Team-based learning enhances long-term retention and critical thinking in an undergraduate microbial physiology course. *Microbiology Education*, 4, 3–12.

*Michaelsen, L. K., Knight, A. B., & Fink L. D. (2002). *Team-based learning: A transformative use of small groups.* Westport, CT: Praeger.

*Michaelsen, L. K., Parmelee D. X., McMahon K. K., & Levine R. E. (2008). *Team-based learning for health professions education: a guide to using small groups for improving learning*. Sterling, VA: Stylus.

Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. *New Directions for Teaching and Learning*, 116, 7–27.

*Nicoll-Senft, J. (2009). Assessing the impact of team-based learning. *Journal on Excellence in College Teaching*, 20, 27–42.

*Nieder, G. L., Parmelee, D. X., Stolfi, A. & Hudes P. D. (2005).Team-based learning in a medical gross anatomy and embryology course. *Clinical Anatomy*, 18, 56–63.

Robinson, D. H., & Walker, J. D. (2008). Technological alternatives to paper-based components of team based learning. *New Directions for Teaching and Learning*, 116, 79–85.

Su, A. Y. (2004). The impact of individual ability, favorable team member scores and student perception of course importance on student preference of team-based learning and grading methods. *Adolescence*, 42, 805–826.

*Shellenberger, S., Seale J. P., Harris, D. L., Johnson, J. A., Dodrill, C. L., & Velasquez M. M. (2009). Applying team-based learning in primary care residency programs to increase patient alcohol screenings and brief intervention. *Academic Medicine*, 84, 340–347.

*Touchet, B. K., & Coon, K. A. (2005). A pilot use of team based learning in psychiatry resident psychodynamic psychotherapy education. *Academic Psychiatry*, 29, 293–296.

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical Statistics with Applications* (7th ed.). New York: Duxbury.

*Weiner H., Plass, H. & Marz, R. (2009). Team-based learning in intensive course format for first-year medical students. *Medical Education*, 50, 69–76.