Constructing and Grading Multiple-Choice Exams:
An Interview with Anthony Marini

by Mike Atkinson, Faculty Associate, Teaching Support Centre

Anthony Marini, 3M Teaching Fellow and professor of Educational Psychology at the University of Calgary, is an expert in measurement and assessment. Recently, we had an opportunity to chat about several issues related to setting and grading multiple-choice exams.

MA: I’m often asked about the number of alternatives one should use for a multiple-choice item. What’s the current thinking on this?

AM: We used to suggest four or five alternatives for an item but recently, the move is toward three well-constructed alternatives.

MA: Are three alternatives really enough?

AM: Absolutely, but they must be well-written, meaningful alternatives. It turns out that five alternatives are not that effective ... it’s simply too hard to construct good distractors (the alternatives that are not correct).

MA: What about the guessing rate ... is 33% acceptable?

AM: Sure, the critical factor is that the distracters should reflect common errors in understanding or reasoning. Too often, the distracters are easily dismissed so the guessing rate is less than you think anyway. You should always run the item analysis and remove those alternatives that are not working.

MA: Before we talk about the item analysis, how many items would you suggest an instructor should use on an exam?

AM: There’s no hard rule about this—you need enough items to generate a valid test. As a rule of thumb, I usually allow 65 seconds per item. Speed should not be an issue and this typically is adequate time for most students.

MA: What about the use of multiple-multiples (e.g., A and B, but never D or E ) ... many people in the professional schools like this kind of item.

AM: The literature is pretty clear on this. Do not use them. The American Medical Association, once of fan of these items, has now abandoned them.

MA: Are they considered too hard?

AM: Initially, multiple-multiples appear harder, but they are essentially an exercise in logical analysis. Once you learn the logic “trick”, the item actually becomes easier and the test loses content validity.

MA: Interesting. Is there a good way to maintain content validity (ensuring that the test accurately reflects the content to be learned) in your test?

AM: The best way is to use a Test Blueprint. Essentially, this is a 2-way table where the rows represent the content actually covered in the course (topics, chapters, etc.) and the columns reflect Bloom’s taxonomy. (http://faculty.washington.edu/krumme/guides/bloom1.html) In this way, you can gage how many items you have included for each topic area and the level of cognitive complexity assessed. Your test should mirror the content actually covered and the weight placed on each of the topics.

Continued on page 2
MA: Let's turn to the issue of item analysis. You've constructed your test, given the exam and graded it. What do we need to do next?

AM: Your multiple-choice test is not complete until you look at the item analysis. No one writes perfect items. You must determine the flaws and correct them.

MA: Where do we start?

AM: Most multiple-choice grading programs will automatically generate an item analysis. Start by looking at the difficulty score for each item. (Editor's note: Scanexam generates a complete set of item analysis statistics). I first delete any item with a difficulty rating of 80% or higher and then re-score the exam.

MA: You delete the most difficult items! Aren't you just increasing the average by getting rid of the hard questions?

AM: Not really. When 80% of the class gets the item wrong, there's probably something wrong with the way the question was phrased. When you look at the item more closely, it's unlikely that only the best students got the item right.

MA: Would you consider re-scoring the item? For example, keep the item with the answer you coded as correct and then accept another alternative as correct also?

AM: No. The item is flawed. Scoring another alternative as "right" does not make the item any better. Delete the item and re-work it for use on another exam.

MA: Would you eliminate the easy items too?

AM: Remember that the goal is to examine the items for flaws. If an item is too easy, change it on the next exam. But the exam is not flawed or unfair simply because an item is easy. You should not penalize the students for your work.

MA: O.K., what's the next step?

AM: Look at the point-biserial correlations and delete any item with a negative correlation. The negative correlation tells you that most of the top students got the item wrong while those who got it right were in the bottom portion of the class. This item does not belong on the exam.

MA: Is there anything else we should look at?

AM: Check the distractor analysis. Evaluate the utility of the distracters. Is anyone choosing this alternative? Who? You want the exam to be a fair test of knowledge and the distracters should draw some attention. If they are not, then change them for the next exam.

AM: We've given the exam, checked the item analysis, deleted the flawed items and made notes for the next exam. Now it's time to turn in the final grades. Let's say that you, or your chair, think that the grades are too high or too low. What's your advice on altering the final grade distribution?

AM: Do not get into these situations. If the final distribution is "too high or too low", the problem is not with the distribution. The assessment instruments (the test) were flawed in some fundamental fashion. If you have been using a blueprint, running item analyses, etc., then you should be able to demonstrate the validity of the exams. Consequently, the final distribution is valid as well. The cure for these problems is to produce valid exams in the first place. Making adjustments to the final distribution is not only poor testing practice, but reinforces the idea that assessment is trivial.

MA: Anything new on the testing horizon?

AM: I'm advocating the use of scoring rubrics for all exams. They are very rich in content and a well-designed rubric can be used over and over.

MA: What exactly is a scoring rubric?

AM: Essentially, it is a scoring guide for questions that not only gives the right answer, but gives examples of an excellent, satisfactory, and an unacceptable answer. These can be shared with students for self-assessment as well as information purposes.

MA: Sounds like rubrics are best designed for essay questions. Can you use them with multiple choice?

AM: Sure. In this case, you would explain why your alternative is correct and why the various alternatives are wrong. This informs both the student and the instructor and keeps us focused on learning.

Note: for information on rubrics, go to: http://www.temple.edu/tlc/handouts/Creating%20Rubrics_Establishing%20Standards.pdf
Enhancing Teaching: Engaging Faculty in the SoTL
Debra Dawson, Director, Teaching Support Centre

In 1990, Ernest Boyer introduced a new term to the academy—the Scholarship of Teaching and Learning (SoTL). In his groundbreaking book Scholarship Reconsidered: Priorities of the Professoriate, Boyer discusses the need to redefine what we mean by scholarship and suggests there are four types of scholarship which occur within the university: the scholarship of discovery (what most of us mean when we talk about research), the scholarship of integration (giving a larger meaning to our research in a more interdisciplinary context), the scholarship of application (applying our research to the world) and finally the scholarship of teaching (where the same rigors of scholarship that are applied to our disciplinary research are also applied to our teaching). Unfortunately Boyer died long before this new form of scholarship could be fully developed. This form of research is becoming more and more common as evidenced by the recent creation of the International Society for the Scholarship of Teaching and Learning (ISSOTL). The 2nd annual conference of this organization was held in the fall of 2005 in Vancouver and attracted over 650 delegates. We know that as faculty start to engage in SoTL this will have a significant impact on how we teach (Huber & Hutchings, 2005). One outcome of Boyer’s work was the CASTL program (Carnegie Academy for the Scholarship of Teaching and Learning) begun by the Carnegie Foundation in 1998. They seek to support SoTL through their Scholars Program and Campus Program (http://www.carnegiefoundation.org/programs/index.asp?key=21).

Although there are several definitions of SoTL, a graphic representation helps to clarify what this form of research is all about. As can be seen in Figure 1, SoTL represents the nexus between scholarly teaching (Kreber & Cranton, 2000), educational research, and traditional disciplinary research. This is a form of practitioner led research into teaching and learning. There is a continuum that exists from scholarly teaching to the scholarship of teaching. Scholarly teaching refers to faculty who consult the literature on ways to improve their teaching and engage in reflective practice about what works and what does not work. They apply much of the same rigor to their teaching that they do to their research. Faculty who engage in SoTL take this one step further by not just performing research on teaching and learning for their own improvement but taking the research further to the point of dissemination in peer reviewed journals and conferences (Richlin, 2001).

Two examples of SoTL are studies that serve to highlight why faculty may wish to perform research in this area. First is the study by Hake (1998) who performed a pre-post analysis of two different methods of teaching introductory physics. Specifically he was trying to determine if there was a better way of changing students’ commonly held beliefs about physics, which interfered with their academic performance at university. A second study by Fullilove and Treisman (1990) examined the differences between those who excel in university mathematics versus those who experience difficulties. Both of these studies provide discipline specific research into teaching and learning and serve to enhance undergraduate education. In a research-intensive university it is incumbent upon us to apply the same lens we do to our teaching that we typically may only have applied to our research. For instance, at Western there are several collaborative research projects being performed to examine the efficacy of using personal response units (“clickers”) in the classroom. As many of these new devices enter our classrooms we can use SoTL to determine when or how they should be implemented.

To learn more about scholarship of teaching and learning, for instance, what journals on teaching research exist, we have included in Reflections an insert written by Maryellen Weimer (editor of The Teaching Professor) who has recently published a book on SoTL or what she calls pedagogic research. We know many of you are interested in determining the best practices in education in your area. We at the Teaching Support Centre welcome the opportunity to assist you with performing this research.

References


New TSC Staff

KIM HOLLAND has joined the Teaching Support Centre as Instructional Designer and Coordinator of Distance Studies. Kim comes to us after 25 years of experience in the Geography Department. Using his experience in teaching and course development using technology, he will assist faculty and graduate students with implementing educational technology in a pedagogically sound manner into face-to-face, hybrid and on-line courses. Distance education course development will be coordinated by Kim who is always interested in talking to Deans, Chairs and course developers in the creation of innovative learning material. Kim will direct workshops, training programs, and seminars on the use of, and the whys of, instructional technologies for both faculty members and graduate students. You can reach Kim at ext. 84612 or e-mail: kholland@uwo.ca.

NADINE LE GROS has joined the Teaching Support Centre as the Language & Communication Instructor. She helps international TAs improve their language skills and communicate effectively in the Canadian classroom. She has 11 years of teaching experience as an ESL instructor and has worked and studied in Europe, the Middle East, and Asia. Nadine works with international TAs in two new programmes this year.

Communication in the Canadian Classroom is a 40-hour series of workshops to improve the spoken and written skills of international teaching assistants. The classes seek to enhance the ITA’s language skills and knowledge of Canadian classroom culture. We examine the educational assumptions of Canadian students, practice giving feedback to students, explore cultural differences in writing, compare linear and circular patterns of communication, and discuss Canadian audience expectations during presentations. The class participants engage in role plays, microteaching sessions, speeches, and letter writing.

Advanced Communication in the Canadian Classroom is a 40-hour series of workshops to improve the ITA’s spoken and written English beyond the basic course. The course was designed for TAs who have been in Canada for an extended period of time and who are fluent in English, but who seek to improve their communication skills. The classes seek to enhance the ITA’s written and spoken communication skills by delivering instruction on differing cultural patterns of reasoning, interpersonal communication, and conflict resolution. We examine the intellectual heritage of the east, which is based on Confucian concepts of balance and harmony, relative to that of the west, which is based on principles of Greek logic and argument. We further examine strategies of persuasion and the effects of cultural predisposition to evidence based on parables, metaphors or physical evidence, and how this is perceived by members of other cultures. We then engage in activities to develop skills to make the ITAs better able to cope in Canada’s thesis-driven academic society and argument-driven interpersonal society. The class participants engage in debates, role plays, and writing tasks such as argumentative essays, letters of application, and research proposals.

If you work with international TAs who may benefit from one of these workshops, please refer them to Nadine at nlegros2@uwo.ca.

Teaching Tip

If you are using WebCT to augment your face-to-face instruction, you have most likely decided to add some version of your lecture notes on the web. Many of you have selected PowerPoint to be the media vehicle to deliver your face-to-face lectures. So the problem is how do I take my PowerPoint lecture and convert it to a form that my students will find useful and I will find easy to do? You have several possible choices.

I have tried several methods and this is what works for my students and me. I take my PowerPoint lecture and convert it to a PDF file. This process is very easy. The software you will need is Acrobat by Adobe (there are several other programs that are available that will do this conversion). Hint: if you are using a Macintosh with OS X you do not need any other software as you can print to a PDF file directly.

You will need to decide within PowerPoint how the student will view the PDF file. I usually select three slides per page as this allows space to the right of each slide image for students to add notes during the class. This PDF file can then be uploaded to WebCT for your students.

Kim Holland
Instructional Designer, TSC
After the Clicks: A PRESSWestern Update
Tom Haffie, Department of Biology; Coordinator, PRESSWestern Project

The PRESSWestern project is dedicated to supporting and evaluating the implementation of personal response units ("clickers") at Western. Clickers are hand-held radio-frequency keypads that are purchased by students in the Bookstore ($50 for two terms) and then registered alongside their student number on the course WebCT site. In the classroom, instructors pose structured questions and then collect, compile, display and record student responses in real time. In contrast to traditional technologies that "push" lectures out to students, clickers "pull" attention and information in from the class. This novel type of "give and take" of ideas in the classroom facilitates various types of active pedagogy that can result in more effective feedback, increased engagement, and improved conceptual learning.

With over 5000 students using clickers this year in over a dozen courses in biology, physics, astronomy, computer science, anatomy, health science, sociology and engineering, Western has taken a leadership role as one of the first and largest campus-wide implementations in the world. An Implementation Group involving instructors and representatives from the Bookstore, Classroom Management Group, ITS, Teaching Support Centre, Centre for New Students, Student Development Centre, SCAAPA, and eInstruction coordinates diverse aspects of the project and provides support for faculty and students through Clicker Clinics and the ITS Help Desk. The Instructional Technology Resource Centre (ITRC) is constructing a website (presswestern.uwo.ca) to provide information and support to the campus community and beyond. As our experience grows and research data accumulate, Best Practices for use of the technology in the classroom are becoming clearer. In particular, we are currently focused on structuring and timing effective questions, testing appropriately and streamlining management of click records.

Dr. John Barnett of the Faculty of Education used an open-ended qualitative online survey to gather student perceptions of the implementation of clickers in Biology 022/023 and Physics 028a. Preliminary analysis of these data revealed that "overall student perceptions were positive, despite the problems of implementation". About 60% of respondents reported experiencing at least one type of technical problem with getting the clickers to work in class. These problems were addressed online and in face-to-face Clicker Clinics sponsored by the Implementation Group. Experience from the first term has reduced such technical difficulties dramatically.

Students identified three main types of advantages to clicker use: attitudinal, interactional and pedagogical. In terms of attitude, students found the clickers to be a fun and convenient mechanism to gain participation credit in a course. The main interactional advantages were engagement with class content and immediate feedback on understanding. Pedagogical advantages identified by students included metacognitive benefits resulting from reconsideration of misconceptions, review of difficult concepts, and awareness of content expectations for upcoming tests. The main disadvantages reported were concerns about the limitations of multiple-choice questions, time wasted on technical issues, potential for cheating during clicker tests, and expense of the clickers themselves.

A second research project is currently underway under the direction of Dr. Debra Dawson (Teaching Support Centre) and Dr. Ken Meadows (Centre for New Students). This project is beginning to characterize the effect of clicker feedback on the help-seeking behaviour of students. Additional post-hoc analyses will also consider correlations between clicker behaviour and overall academic success as well as relationships between the cognitive level of clicker questions relative to exam questions.

Favourable media coverage and the academic "grapevine" have already established Western as a leader in clicker technology. Now that the heavy lifting of technical implementation is behind us, we will focus on refining the use of clickers, supporting the expansion of this project on campus and sharing our expertise with others.

For further information, contact Tom Haffie at thaffie@uwo.ca.

The PRESSWestern project acknowledges significant support in cash and/or in kind from the Faculty of Science, eInstruction, a UWO Fellowship in Teaching Innovation Award, Pearson Education, the Department of Biology, the Department of Physics and Astronomy, the Faculty of Medicine and Dentistry, the Centre for New Students, the Student Development Centre, the Teaching Support Centre, and Information and Technology Services.
The Teaching Support Centre at The University of Western Ontario has a site license allowing the University Community free access to *The Teaching Professor*, a higher education newsletter produced by Magna Publications. *The Teaching Professor* is a valuable resource featuring innovative ideas to enhance your teaching.

Edited by respected scholar and expert Dr. Maryellen Weimer of Penn State Berks Lehigh, *The Teaching Professor* is a forum for discussion of the best strategies supported by the latest research for effective teaching in the college classroom. From tips for class discussion to mentoring fellow faculty, *The Teaching Professor* stretches from the theoretical to the highly specific. Typical topics include assessment and evaluation, engagement of student interest, faculty time management, and the learner-centered classroom.

You can access this publication by going to our website: [www.uwo.ca/tsc](http://www.uwo.ca/tsc). If logging in from a campus computer, you will not need to enter a password. To access *The Teaching Professor* from your home computer, you will first need to subscribe (it is free) by going to: [www.magnapubs.com/licenses/subscribe.html](http://www.magnapubs.com/licenses/subscribe.html) and login as:

Voucher Code: UWOCA       PIN Number: 2903

After you subscribe, you will receive an e-mail directing you to the most current newsletter and the archives. **Note: you must subscribe if you wish to be notified by e-mail when each new issue is available.**

---

The following article is taken from *The Teaching Professor* (March 2005) and is an example of the growing area of Inquiry-Based Learning. Hudspith and Jenkins define Inquiry as “a self-directed, question-driven search for understanding”(p.9). Inquiry is student centered, born out of the curiosity, questions and uncertainties of the student which can lead to discovery and understanding. The aim of Inquiry-Based Learning is to develop, in the student, the skills to be a lifelong learner. After all, that is one of the goals that we all hope our students will achieve.


**Peer-Led Team Learning, Fewer Lectures: More Learning**

(Reprinted with permission from Magna Publications)

Faculty reluctance to use student-centered approaches often stems from the fear that with less content being covered, less learning will occur. Some empirical studies addressing that issue have results some will find surprising.

In a section of a large general chemistry course, faculty experimented with a Peer-Led Team Learning (PLTL) method developed by the National Science Foundation. This approach puts students in groups of 10 and assigns each group a peer leader who is a fellow undergraduate student, but one who has successfully completed the course. In this study, student groups met for one 50-minute PLTL session per week. This session replaced one of the three 50-minute lectures scheduled for each week. During the PLTL sessions, students worked through one or two activities in which they explored “information in order to discover the need for new (to the students) concepts and subsequently ‘invent’ and apply those concepts...” (p. 135) Content in the activities preceded lecture material on the topic. Peer leaders (who were trained) acted as facilitators.

The control group in the study consisted of a second section of the same general chemistry course. It was taught by the same instructor, included the same content, and was graded similarly with 75 percent of the student’s grade being determined by performance on course exams and a final. Tests for both sections were identical and were given at the same time during the semester.

Continued on page 7
Because students could not be randomly assigned to these sections, researchers ran a number of tests to compare student populations in the two sections. None of the differences they tested for were significant, leaving them to conclude that the two students groups were substantially the same.

Performance on course exams and the final were used to ascertain the effectiveness of the PLTL activity. And here’s what the researchers found: “the experimental group consistently outperformed the control group on the course exams and final exam.” (p. 136) All the reported differences were at statistically significant levels. Moreover, another statistical analysis revealed that “the differences in performance between the two sections became larger as the course progressed.” (p. 136) And, a regression analysis led to this conclusion: “a student who attends PLTL sessions can be expected to perform better on exams than another student at the same SAT level. This is especially impressive considering students in the PLTL sessions did receive one less lecture per week than those in the control group.” (p. 138)

Student survey data showed that students thought that the PLTL experiences were instrumental in their learning and performance. Seventy-four percent said that the sessions were beneficial. Seventy-six percent thought the PLTL sessions made up for the missed lecture. Eighty-five percent reported that if given the chance they would continue to participate in the PLTL sessions.

Next to be tested, the researchers note, is the impact of an intervention like this on long-term retention of content. But they conclude with this bottom line: “Fears that students who had less exposure to lecture would learn less proved to be groundless in this study.” (p. 139)


Copyright © 2006 Magna Publications, Inc. All rights Reserved.
Different Patterns of Communication
Nadine Le Gros, Language & Communication Instructor, Teaching Support Centre

Have you ever asked a student, “Do you understand?” only to have the student answer in the affirmative when you just knew that he or she did not understand? Many factors will result in this answer, most of which involve the issue of saving face. This situation is exacerbated with international students because of cross-cultural patterns of communication. Only, did you know that this white lie was actually about trying to save you face?

In Canada, responsibility for understanding is placed on the speaker’s shoulders. This is why we explain things step by step: we want to be very clear, and we don’t want to forget to include any points. We don’t mind it when people tell us that they don’t understand; in fact, we expect our students to do so. If they do not, we might feel stymied while we are trying to teach.

In many other countries, especially those in Asia and the Middle East, it is the listener’s responsibility to understand what the speaker is saying. Speakers will imply a great deal, and it is the listener’s responsibility to infer exactly what the speaker is saying. In countries where this is a communication pattern, students will not tell an instructor that they do not understand: to do so would be extremely rude and would constitute a loss of face for the instructor.

So, how can we overcome this difference in how we communicate in order to ensure good teaching and good learning? I have three suggestions:

• At the end of a class, ask the students to take one minute to write down what the muddiest point for them is on index cards. They can submit these anonymously or they can include their names.

• Ask students open-ended concept questions, and generally try to avoid questions that require a yes/no answer. For example, if you were discussing the issue of academic honesty, concept questions would involve asking for examples of plagiarism.

• Explain this difference in communication patterns, and reassure your students that you want and need to know when they don’t understand. Do, however, understand that mere knowledge of this pattern won’t be enough for the students to overcome the situation immediately. You will need to teach them that it’s okay to tell you when they don’t understand.