Grading Scheme for Veterinary Student Performance in Pass–Fail Didactic Surgery

F. A. Mann,* Zachary March, James L. Tomlinson, Eric R. Pope, James L. Cook, Colette C. Wagner-Mann, and Hun-Young Yoon

ABSTRACT A retrospective study was performed to evaluate a satisfactory/unsatisfactory (S/U) grading scheme in a didactic surgery laboratory during the first 3 years of implementation (2002–2004) and identify areas for improvement that might be adapted to this course or similar courses. Each instructor graded six students per session by assigning a descriptor of very good (G), acceptable (A), or unacceptable (U) in each of 11 assessment categories. A U in any category in one of the final two laboratories resulted in a failing grade for the course, unless the student performed acceptably in a make-up laboratory. A computerized course evaluation was used to solicit student feedback. Also, the numbers of G, A, and U grades were used to evaluate consistency of grading among instructors, to compare resident and faculty scores, and to track student progress. The return of course evaluations was low, but those available indicated favorable acceptance of S/U grading. There was little difference in assigned grades between faculty and residents, but some individual instructors seemingly graded more strictly than others. Student grades improved as the course progressed each year. No student received a final failing grade; however, two students required the make-up laboratory. Efforts to improve subjective grading should include planned acquisition of student feedback and establishment of more consistency of grading. While objective criteria may not be enough to adequately assess overall performance in didactic surgery laboratories, consistency of subjective evaluation requires adherence to well-defined assessment criteria.

articles

Efforts to reduce laboratory animal usage in teaching surgery to veterinary students have resulted in some effective instructional models (Greenfield et al., 1993, 1995; Griffon et al., 2000; Olsen et al., 1996; Smeak et al., 1994). Further, these models have provided avenues for assessing specific surgical skills. However, the importance of instruction using live animals is still recognized (Greenfield et al., 1995; Hedlund et al., 2002), and surgeons charged with teaching surgery laboratories must assess overall student performance and record grades according to the specifications of their respective institutions.

In 2001, the didactic surgery laboratory of this report became a stand-alone course requiring its own grading scheme. Before 2001, the laboratory was part of a lecture course, and the grades (A, B, C, D, or F) assigned to the students were calculated from scores on written examinations based on lecture material (see Appendix). The surgical faculty did not feel that written examinations were appropriate for assessing surgical skills; therefore, they considered other means of grading laboratory performance. A pass–fail system was considered because it would be conducive to subjective evaluation of surgical skills. The downside to a pass–fail system would be the inability to differentiate among the passing students to record who was excellent and above average. Further, the pass–fail nature of the final grade would preclude any contribution of the final grade to a student's overall grade point average. Such lack of differentiation could hinder selection of students for scholarships, awards, and advanced training programs.

Medical schools seem to favor letter grading or like systems to differentiate among students in surgical training (Dietrick et al., 1991; Moss et al., 1978; Ravelli and Wolfson, 1999; Reznick et al., 1989). The authors assume

Impact Statement

Subjective assessment of skills in a laboratory setting is challenging, especially when there are a large number of students and a variety of instructors doing the grading. This manuscript describes a novel method of performance assessment and assignment of grades utilizing a rubric-driven grading scheme that achieves prompt feedback and attempts to minimize grader variability. Experience with this grading method during the first 3 years of implementation is reported.

F.A. Mann, J.L. Tomlinson, E.R. Pope, J.L. Cook, and H.-Y. Yoon, Dep. of Veterinary Medicine and Surgery, College of Veterinary Medicine, Univ. of Missouri, 900 East Campus Drive, Columbia, MO 65211; Z. March, Office of Information Technology, College of Veterinary Medicine, Univ. of Missouri, 900 East Campus Drive, Columbia, MO 65211; C.C. Wagner-Mann, Dep. of Surgery, School of Medicine, Univ. of Missouri, 900 East Campus Drive, Columbia, MO 65211. (Dr. Wagner-Mann's current affiliation is Helias High School, Jefferson City, MO. Dr. Pope's current affiliation is Ross University School of Veterinary Medicine, St. Kitts. Dr. Yoon's current affiliation is the Department of Veterinary Surgery, College of Veterinary Medicine, University of Konkuk, Seoul, South Korea.) Received 21 May 2007. *Corresponding author (MannF@missouri.edu).

J. Nat. Resour. Life Sci. Educ. 38:61–70 (2009).
http://www.JNRLSE.org

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: A, acceptable; G, very good; PDA, personalized data assistant; S/U, satisfactory/unsatisfactory; U, unacceptable.

that this preference is prevalent among veterinary schools, but could find no reports to substantiate that assumption. Recognizing the pros and cons of pass–fail grading schemes, the surgical faculty embarked on modifications in laboratory instruction that would allow critical assessment of students' surgical performance with the final grade of satisfactory or unsatisfactory according to university guidelines. That didactic surgery laboratory grading scheme is the subject of this report. The investigational objectives of this study were to describe this grading scheme, demonstrate its effective use during the first 3 years of its implementation, and identify areas for improvement that might be adapted to this grading scheme or to similar courses taught at other institutions.

Materials and Methods

The surgery laboratory course was offered in an 8-week instructional period. There were two class meetings during the first week of the course aimed at orientation and organization. The first class period was an orientation session in a lecture setting during which logistics of the course were outlined and the grading mechanism was explained. The second class period was an observational laboratory with three stations: (1) prepping of the patient; (2) scrubbing, gowning, gloving, and quarter draping; and (3) wrapping surgical packs. Instructors at these stations demonstrated the techniques that were to be expected in the laboratory. Videotapes and computer video clips of these procedures were available for the students to review. Additional orientation was not necessary because the students had experience with instrument identification and usage, suturing techniques including hand ties, and basic psychomotor skills using models, such as an abdominal surrogate made of cloth fabric (Dog Abdominal Surrogate for Instructional Exercises—DASIE; DASIE International, Elora, ON, Canada N0B 1S0), in laboratories during the soft tissue surgery course taught in the immediately prior 8-week instructional period.

The first laboratory during the second week of the block was designed for the students to show proficiency in the techniques demonstrated to them the week before. Without use of animals, the first laboratory consisted of surgical pack preparation, scrubbing, gowning, gloving, draping, and suturing an inanimate object while in full surgical attire. The remaining six laboratories utilized live dogs for soft tissue surgery laboratories (n = 3) and cadavers (dogs euthanatized from the week before) for the orthopedic laboratories (n = 3). Ophthalmic surgical procedures were performed with ophthalmology faculty instructing at the end of one of the orthopedic laboratories, but ophthalmologists did not contribute to grading except for contributing quiz questions in 2004. The surgical procedures are listed in Table 1. Use of these dogs in teaching was approved by the University Animal Care and Use Committee and consistent with the United States National Institutes of Health *Guide for the Care and Use of Laboratory Animals* and the Animal Welfare Acts. Dogs were purchased through the University Office of Animal Resources.

Students were divided into pairs based on height to accommodate appropriate surgical table height for all students, and were assigned to 1 of 12 surgical cubicles on Tuesday, Wednesday, or Thursday. Each student manned the same cubicle the same day each week. Additionally, each student was assigned two anesthesia days, one student for each dog in the first and last live-animal laboratories, and two students for each dog in the second live-animal laboratory, which included diaphragmatic herniorrhaphy. Assigning two students per dog in the diaphragmatic hernia laboratory ensured that each student experienced manually ventilating a dog with an open thoracic cavity. Faculty anesthesiologists directed and taught the anesthesia experience, but they did not grade the students for anesthesia performance.

Surgical faculty and residents served as laboratory instructors for the surgical procedures. One instructor was assigned to three cubicles (six students) per laboratory. An attempt was made to have each faculty surgeon assigned to each of the cubicles before the end of the course, but sometimes scheduling conflicts precluded such. Instructor assignments are outlined in Table 2. Instructors were

Table 1. Schedule of orientation, laboratories, quizzes and final examination for a didactic surgery laboratory (2004 dates).

Date	Activity
5 May	Orientation part 1 (introductions and overview)
6 May	Orientation part 2 (demonstration laboratory)
10 May	Quiz over 11–13 May laboratory (10 points)
11–13 May	Laboratory 1 (pack preparation, scrubbing, gowning, gloving, draping)
17 May	Quiz over 18–20 May laboratory (10 points)
18–20 May	Laboratory 2 (celiotomy 1: exploratory celiotomy, ovariohysterectomy, kidney and lymph node biopsy)
24 May	Quiz over 25–27 May laboratory (10 points)
25–27 May	Laboratory 3 (orthopedics: intramedullary pinning)
28 May	Quiz over 1–3 June laboratory (10 points)
1–3 June	Laboratory 4 (celiotomy 2: gastrotomy, diaphragmatic herniorrhaphy, liver biopsy, thoracostomy tube placement)
7 June	Quiz over 8–10 June laboratory (10 points)
8–10 June	Laboratory 5 (orthopedics: external fixator application; and soft tissue surgery: castration)
14 June	Quiz over 15–17 June laboratory (10 points)
15–17 June	Laboratory 6 (celiotomy 3: ovariohysterectomy via limited approach, intestinal resection and anastomosis, intestinal biopsy, cystotomy)
21 June	Quiz over 22–24 June laboratory (10 points)
22–24 June	Laboratory 7 (1) orthopedics: cranial cruciate ligament stabilization; (2) Ophthalmic procedures
25 June	Final examination—comprehensive (30 points)
29 June	Make-up laboratory†

† The make-up laboratory was reserved for students with unacceptable marks in the final two laboratories and students with an excused absence from one laboratory. There were no make-up laboratories for unexcused absences. An unexcused absence would result in a failing grade for the course. An incomplete grade could be given for missing more than one laboratory (or failing to make up the missed laboratory) if there was official documentation of hardship.

Table 2. Instructor assignments for a didactic surgery laboratory (2004 dates).†

Lab no.	Date	Day	Tables 1–3	Tables 4–6	Tables 7–9	Tables 10–12	Procedures
1	11 May	Tues.	Fac F	Fac A	Fac C	Res C	pack preparation
	12 May	Wed.	Fac F	Fac A	Fac C	Res C	scrubbing, gowning
	13 May	Thur.	Fac F	Fac A	Fac C	Res C	gloving, draping
2	18 May	Tues.	Fac A	Fac F	Fac B	Res D	
	19 May	Wed.	Fac A	Fac F	Fac B	Res D	celiotomy 1
	20 May	Thur.	Fac A	Fac F	Fac B	Res D	
3	25 May	Tues.	Fac C	Fac E	Fac F	Res E	
	26 May	Wed.	Fac C	Fac E	Fac F	Res E	orthopedic 1
	27 May	Thur.	Fac C	Fac E	Fac F	Res E	
4	1 June	Tues.	Res F	Fac B	Fac A	Fac F	
	2 June	Wed.	Res F	Fac B	Fac A	Fac F	celiotomy 2
	3 June	Thur.	Res F	Fac B	Fac A	Fac F	
5	8 June	Tues.	Fac F	Res F	Res C	Fac C	
	9 June	Wed.	Fac F	Res F	Res C	Fac C	orthopedic 2 also: castration
	10 June	Thur.	Fac F	Res F	Res C	Fac C	
6	15 June	Tues.	Fac B	Fac F	Res E	Fac A	
	16 June	Wed.	Fac B	Fac F	Res E	Fac A	celiotomy 3
	17 June	Thur.	Fac B	Fac F	Res E	Fac A	
7	22 June	Tues.	Res D	Fac C	Fac D	Fac F	
	23 June	Wed.	Res D	Fac C	Fac D	Fac F	1. orthopedic 3 2. ophthalmic‡
	24 June	Thur.	Res D	Fac C	Fac D	Fac F	

† Fac, faculty surgeon; Res, surgery resident.
‡ Ophthalmology clinicians (two or three faculty and one or two residents) circulated among the tables in the second half of Laboratory 7, but were not assigned to specific tables.

oriented regarding the grading mechanism with emphasis on consistency of grading. Likewise, instructors were told to avoid deviation from the written descriptions of the laboratory exercises in the course notes so as to avoid confusion among the students.

Grading was conducted in two formats: (1) objective and (2) subjective. Students were required to pass both objective and subjective grading assessments to pass the course with an S grade. Objective grading consisted of seven 10-point pre-laboratory quizzes and one 30-point laboratory final examination. Students were required to score at least 77 out of these 100 points to pass the objective portion of the grading. This pass point was chosen to encourage thorough laboratory preparation while allowing an opportunity to offset poor quiz performance with a high grade on the final examination. The quizzes were administered in a computer laboratory on the Monday of each week (Friday of the week before when necessary to accommodate for a national holiday). The final examination was administered in a computerized format (Toolbook Instructor, SumTotal, Bellevue, WA) on the last Friday of the block. The timing of quizzes and final examination relative to the surgical procedures is illustrated in Table 1. Students had immediate access to their computerized quiz and final examination scores as soon as they locked their answers and submitted their examinations.

The subjective portion of the grading was more complex than the objective portion and required close cooperation among instructors and the information technology staff who maintained the scoring spreadsheets. A computer station was set up in the teaching laboratory for entry of grades by instructors before they left the laboratory for the day. A secure location on the computer contained a spreadsheet (Microsoft Excel, Microsoft Corporation, Redmond, WA) personalized to the instructors of the day and easily accessible to them via a shortcut once they logged onto the computer. The spreadsheet contained 11 grading categories to assess for each student, and instructors were given written guidelines in the form of questions for each category to enhance consistency of grading (Table 3). Students were also provided these guidelines in their course notes so that they would be informed as to what was being evaluated as they performed the laboratory exercises.

Information technology personnel emailed personalized copies of the score sheets for each instructor as a printable electronic file (Adobe Acrobat, Adobe, San Jose, CA) to each instructor at the beginning of the instructor's assigned week in the laboratory. The instructor printed the personalized score sheet (Table 4) and took it to the laboratory for making notes during the laboratory. [Note: In 2002, the personalized spreadsheet was provided in PDA format using Palm Pilots (Palm Pilot, Palm Corporation, Sunnyvale, CA) issued to the laboratory instructors, who would download their entries into the laboratory computer at the end of each laboratory. The PDAs were found to be cumber-

Table 3. Didactic surgery laboratory grading categories and guidelines for instructors.

Grading category	Guideline questions
Patient preparation	• Was the patient clipped appropriately for the intended procedures?
	• Was the animal positioned appropriately?
	• Did the student correctly perform open gloving?
	• Was the rough scrub performed correctly?
	• Was the sterile scrub performed correctly?
	• Were the student surgeons appropriately attired for the sterile scrub?
Scrubbing	• Did the student demonstrate knowledge of scrubbing oneself in preparation for surgery?
	• Was the student appropriately attired for the scrubbing procedure?
	• Did the student scrub in proper order from fingers to elbows?
	• Were there any breaks in aseptic technique while scrubbing?
	• If a break in technique occurred, did the student appropriately correct for the error?
Gowning and gloving	• Did the student demonstrate knowledge of gowning and gloving oneself (closed gloving technique) in preparation for surgery?
	• Was the student appropriately attired for the gowning and gloving procedure?
	• Did the student gown at the appropriate time?
	• Did the student don gloves at the appropriate time?
	• Were there any breaks in aseptic technique while gowning or gloving?
	• If a break in technique occurred, did the student appropriately identify and correct for the error?
	• Did the student try to secure the gown's sterile back prematurely (i.e., before gloving)?
Pack preparation and pack opening	• Were the packs appropriately labeled?
	• Were the packs tidy?
	• Once opened, did the packs contain the appropriate instruments and other materials? [Note: Students are not penalized for erroneous packing of the instrument pack because this pack is prepared by students other than those using the instrument pack.]
	• Did the student open packs at the correct time?
	• Were the packs opened in appropriate locations?
	• Were the packs opened without a break in aseptic technique?
	• If a break in technique occurred, did the student appropriately identify and correct for the error?

(continued on next page)

some by the instructors and were discontinued in 2003.] At the end of each laboratory, the instructor logged onto the computer and changed scores on his/her spreadsheet as appropriate. If the instructor chose to give all A grades, there was no need to log on, because A was the default score. The instructor had the option of changing the A to G if above average performance was noted or to U if the student's performance was not acceptable. Whenever a U grade was entered there was a prompt for mandatory text for the instructor to comment about the student's performance. Text could also be typed to offer comments for G and A grades, but was only required for U grades. Student received their grades, including comments as appropriate, via email the morning following each laboratory.

Instructors were asked to grade strictly in the first five laboratories, because these were the laboratories where the students were developing their skills and a U grade would not affect the final grade. Strict grading consisted of giving a U grade for even a minor error in any category regardless of whether the error was noted and corrected by the student during the laboratory. In the final two laboratories the instructors were told to assign a U grade when the error would ordinarily cause significant patient morbidity or severely damage surgical instrumentation. A U grade in any category in either of the final two laboratories would result in a failing grade for the course. An afternoon was set aside on the final day of the instructional period for a make-up laboratory where a student who made a U in one of the final two laboratories would have an opportunity to repeat the laboratory (or portion thereof) to demonstrate proficiency. The appropriate (soft tissue or orthopedic) faculty surgeon (not a surgical resident) would conduct the make-up laboratory. The make-up laboratory time was also available for students who legitimately missed a laboratory due to hardship.

Computerized course evaluations were automatically sent to students via email 2 weeks upon completion of the course followed by a reminder 6 weeks later, according to policies established for all courses in the college of veterinary medicine curriculum. The course evaluation contained 30 statements, six of which were related to testing or grading. For each question the students chose one of the following: 4 (strongly agree), 3 (moderately agree), 2 (moderately disagree), 1 (strongly disagree), or NA (not applicable). Students also had the opportunity to provide free text answers to the following two questions:

1. "What aspects of the teaching or content of this course do you feel were especially good?"

Table 3, continued

Table management	• Was the Mayo stand positioned appropriately at the end of the surgery table?
	• Did the student keep an orderly table?
	• Were frequently used instruments kept nearer to the surgical field than less frequently used instruments?
	• Were unused instruments kept in safe locations (i.e., locations that would prevent personal injury and inadvertent loss of instruments from the table)?
	• Were instruments kept clean of excess blood?
	• For body cavity surgeries, was a sponge count performed before making the skin incision and immediately before closing the body cavity?
Aseptic technique	• Did the student demonstrate knowledge of aseptic technique during preparation of the patient, during preparation of oneself for surgery, and during the surgical procedure?
	• Was the student respectful of sterile fields?
	• Did the student anticipate potential breaks in aseptic technique and take necessary measures to avoid breaches of sterility?
	• If a break in aseptic technique occurred, did the student appropriately identify and correct for the error?
Instrument handling	• Did the student handle all instruments properly?
	• Did the student use the correct scissors for cutting tissue and for cutting sutures?
	• Did the student safely load the scalpel blade onto the handle?
	• Did the student apply hemostatic forceps appropriately to bleeding vessels and to vascular pedicles?
Suturing	• Did the student use appropriate suture material and suture patterns for the techniques performed?
	• Did the student avoid handling the needle while it was embedding in tissue?
	• Did the student tie square knots correctly via instrument and, when appropriate, hand ties?
	• Did the student know when to use (and not use) a surgeon's knot?
	• Did the student properly tie ligatures?
	• Did the student understand and demonstrate how to accurately appose various layers during closure?
	• Did the student space sutures and suture passages appropriately?
	• Did the student cut suture tags to the appropriate length?
Tissue handling	• Did the student show respect for tissues?
	• Were the tissues kept moist?
	• Were instruments applied to tissues in delicate fashion?
	• Were there many wasted intra-operative movements and maneuvers?
Knowledge of anatomy	• Was it evident that the student had reviewed pertinent anatomy?
	• Did the student apply knowledge of anatomy to make intra-operative decisions?
Knowledge of the procedure	• Was it evident that the student had prepared for the procedure by reading the notes and utilizing other resources?
	• Did the student know the proper order of procedures?
	• Did the student understand why procedures were performed in the manners described in the notes?
	• Did the student perform the procedures properly?

2. "What changes could be made to improve the teaching or the content of this course?"

Students had 1 week after the reminder to complete the evaluations. Then, information technology staff collated the results and forwarded them to the instructional leader via email.

The number of U grades was recorded for each category and each laboratory to see where most errors were made. The number of G, A, and U grades assigned by residents and faculty were compared using Chi Square Analysis to see if one group was apparently stricter than another in grading. Statistical significance was assumed at $p < 0.05$. Individual faculty and resident grade assignments were also evaluated to identify individual instructors who might be stricter graders than others. The numbers of G, A, and U grades in laboratories 1, 5, and 7 were compared to evaluate overall improvement or deterioration of student performance over the course of the laboratory each year.

Results

There were 193 students (2002: $n = 62$; 2003: $n = 64$; 2004: $n = 67$) graded according to the S/U grading mechanism described in this study. No student failed the course; however, in 2002, two students each made one U in the "Knowledge of procedure" category in the final soft tissue surgery laboratory. A faculty soft tissue surgeon instructed and evaluated those two students in the make-up laboratory, which both students passed; therefore, both students passed the course.

There were 31 course evaluations returned from students (2002: $n = 18$, a 29.0% response; 2003: $n = 8$, a 12.5% response; 2004: $n = 5$, a 7.5% response). The average response rates for these students in their other college of veterinary medicine required courses were 28.3, 11.7, and 8.8% in 2002, 2003, and 2004, respectively. Table 5 illustrates scores for evaluation statements that dealt with

Table 4. Personalized score sheet for didactic surgery laboratory grading. [Note: Before the laboratory each week, each instructor received via email a copy of this form with the date, laboratory instructor's name, table numbers, and student names completed.]†

VMS 607B Didactic Surgery Laboratory Subjective Grading Form

Student name:	**Table #**_____		**Table #**_____		**Table #**_____	
Patient preparation	A	A	A	A	A	A
Scrubbing	A	A	A	A	A	A
Gowning and gloving	A	A	A	A	A	A
Pack opening/ Pack preparation	A	A	A	A	A	A
Table management	A	A	A	A	A	A
Aseptic technique	A	A	A	A	A	A
Instrument handling	A	A	A	A	A	A
Suturing	A	A	A	A	A	A
Tissue handling	A	A	A	A	A	A
Knowledge of anatomy	A	A	A	A	A	A
Knowledge of procedure	A	A	A	A	A	A
COMMENTS						

† Grades of G = very good; A = acceptable; U = unacceptable (comments required with U).

testing and grading in the didactic laboratory. Free text comments included three statements (2002: *n* = 1; 2003: *n* = 2) indicating that the pre-laboratory quizzes were especially good and two statements (2002: *n* = 1; 2003: *n* = 1) indicating that the pre-laboratory quizzes could be improved. There was one comment in 2002 expressing favor for the course being graded in a pass–fail fashion instead of using traditional letter (A–F) grading. In 2004, there were two grading-related comments, one comment indicating some inconsistency among instructors regarding the assignment of Us, and one comment that some instructors gave more feedback on performance than others.

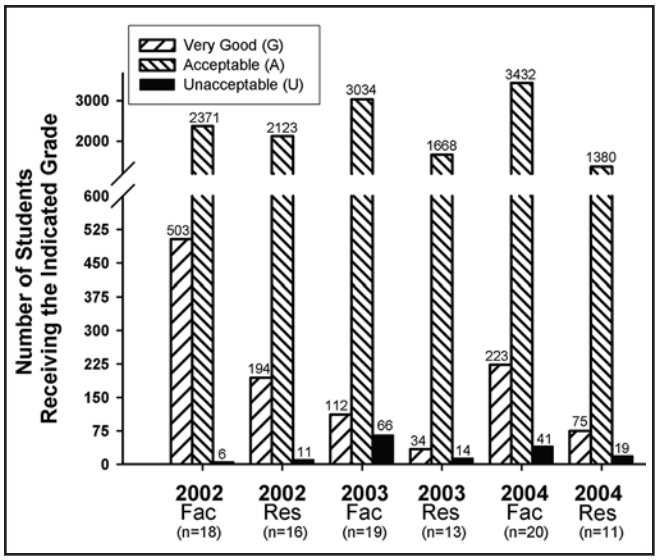

Fig. 1. The numbers of very good (G), acceptable (A), and unacceptable (U) grades assigned by faculty surgeons (Fac) and surgical residents (Res) during didactic surgery laboratories in 2002, 2003, and 2004.

There were 838 Gs, 8837 As, and 113 Us assigned at 57 laboratory stations (two students per station) graded by faculty surgeons, and there were 303 Gs, 5171 As, and 44 Us assigned at 40 laboratory stations (two students per station) graded by surgery residents (Fig. 1). The categories with the most Gs assigned by faculty were pack preparation (*n* = 98) in 2002, tissue handling (*n* = 34) in 2003, and knowledge of procedure (*n* = 37) in 2004. The category with the most Gs assigned by residents was knowledge of procedure (*n* = 37 in 2002; *n* = 17 in 2003; and *n* = 24 in 2004). The categories with the most Us assigned by faculty were pack preparation and knowledge of procedure (*n* = 2, each) in 2002, aseptic technique (*n* = 17) in 2003, and aseptic technique and instrument handling (*n* = 12, each) in 2004. The categories with the most Us assigned by residents were knowledge of procedure (*n* = 3) in 2002, instrument handling (*n* = 6) in 2003, and aseptic technique, instrument handling, and tissue handling (*n* = 4, each) in 2004.

More Gs were assigned in 2002 than in 2003 or 2004 ($p = 0.026$), but overall there was no significant difference in the number of Gs assigned by faculty compared with the number of Gs assigned by residents ($p = 0.566$). Also, there was no significant difference in the number of As assigned by faculty compared to the number of As assigned by residents ($p = 0.273$). Overall, faculty assigned more Us than residents ($p = 0.029$). In 2002, residents assigned more Us than did faculty, and in 2003, faculty assigned more Us than did residents, but in 2004 there was no difference in the number of Us assigned.

The numbers of Gs, As, and Us per student assigned by each instructor for all 3 years combined are listed in Table 6. There was no significant difference between faculty and residents in the numbers of Gs ($p = 0.566$), As ($p = 0.863$), and Us ($p = 0.310$) assigned per student. However, when all faculty grades combined were compared with all resident grades combined, faculty were more likely to

Table 5. Scores from course evaluations submitted by students after a didactic surgery laboratory course in 2002, 2003, and 2004.

Statement	Year	Score, %					No. of responses	Avg. score
		4†	3	2	1	NA		
Objective examinations were given as often as needed.‡	2002	22.22	5.56	38.89	27.78	5.56	18	2.24
	2003	87.5	12.5	0	0	0	8	3.88
	2004	80	20	0	0	0	5	3.8
Prompt useful feedback on exams was provided.	2002	61.11	27.78	5.56	0	5.56	18	3.53
	2003	50	12.5	25	12.5	0	8	3.00
	2004	80	20	0	0	0	5	3.8
The exams concentrated on the more important aspects of the course.	2002	55.56	33.33	5.56	0	5.56	18	3.53
	2003	62.5	12.5	25	0	0	8	3.38
	2004	80	20	0	0	0	5	3.8
The exams were clearly worded.	2002	50	33.33	5.56	0	11.11	18	3.50
	2003	50	25	12.5	12.5	0	8	3.13
	2004	60	20	0	0	20	5	3.75
My grades accurately reflect my performance in the course.	2002	66.67	22.22	5.56	0	5.56	18	3.65
	2003	62.5	25	0	12.5	0	8	3.38
	2004	80	20	0	0	0	5	3.8
The grading system was adequately explained and fair.	2002	55.56	38.89	0	0	5.56	18	3.59
	2003	50	37.5	0	12.5	0	8	3.25
	2004	80	20	0	0	0	5	3.8

† 4 = strongly agree; 3 = moderately agree; 2 = moderately disagree; 1 = strongly disagree; NA = not applicable

‡ The actual wording for the first statement in 2002 was "Examinations should be given more frequently."

Table 6. The numbers of very good (G), acceptable (A), and unacceptable (U) grades per student assigned by each laboratory instructor from 2002 through 2004.†

Instructor	No. of students‡	Gs per student	As per student	Us per student
Fac A	176	0.6	10.1	0.6
Fac B	135	0.9	10.4	0.0
Fac C	124	0.6	10.4	0.1
Fac D	95	2.1	9.5	0.0
Fac E	18	0.0	11.0	0.0
Fac F	320	1.1	12.7	0.0
Res A	80	0.3	11.6	0.1
Res B	91	0.7	10.7	0.0
Res C	106	0.4	10.9	0.1
Res D	100	1.1	10.1	0.1
Res E	71	0.0	10.9	0.1
Res F	36	1.5	9.2	0.4

† Fac vs. Res: numbers of Gs ($p = 0.566$), As ($p = 0.863$), and Us ($p = 0.310$) assigned per student. Combined Fac grades vs. combined Res grades: number of Gs ($p < 0.001$) and number of Us ($p = 0.769$). Fac D and Res F vs. all other graders: number of Gs ($p = 0.001$). Fac A and Res F vs. all other graders: number of Us ($p = 0.002$). Fac = faculty surgeon; Res = surgery resident.

‡ The number of students indicates the total number students who received a grade from the instructor listed. If an instructor graded a student more than once because of being assigned to that student's table in more than one laboratory, then that student counted as more than one student in that instructor's number of students.

assign a G than were residents ($p < 0.001$). When it came to assigning Us, there was no difference between faculty and residents (1% of all grades assigned were Us for both faculty and residents; $p = 0.769$). The data in Table 6 illustrates that some instructors (faculty surgeons D and F and surgery residents D and F) tended to give more Gs than other instructors, but only faculty surgeon D and surgery resident F gave significantly more Gs than the other graders ($p = 0.001$). Two instructors (faculty surgeon A and surgery resident F) were more likely to give Us ($p = 0.002$). One instructor (faculty surgeon A) stands out as being balanced in the number of Gs and Us assigned.

The number of Gs, As, and Us assigned in Laboratories 1, 5, and 7 are recorded in Table 7. The number of Gs increased progressively from Laboratory 1 to Laboratory 5 to Laboratory 7 in 2002 and 2004, but not in 2003, when there were no Gs given in Laboratory 5. However, there were seven times more Gs in Laboratory 7 compared with Laboratory 1 in 2003. The number of Us in Laboratory 5 was the same as in Laboratory 1 in 2002, lesser than Laboratory 1 in 2003, and greater than Laboratory 1 in 2004. No Us were assigned in Laboratory 7 in any of the three years.

Discussion

The low student response rate on the computerized course evaluation made it difficult to use student feedback to assess the grading mechanism of this course; however, the input received gave the grading mechanism generally positive marks. The mechanism employed by the college

Table 7. The total number of very good (G), acceptable (A), and unacceptable (U) grades assigned to third-year veterinary students in Laboratories 1, 5, and 7 of a didactic surgery laboratory in 2002 through 2004.

Year	Lab no.	G	A	U
2002	1	64	676	4
	5	105	635	4
	7	198	546	0
2003	1	6	680	18
	5	0	704	0
	7	42	662	0
2004	1	0	737	0
	5	11	718	8
	7	97	640	0

administration during the time frame of this study was not conducive to a high student response rate because the call for input came while students who took the course were on an 8-week summer break. Despite this timing disadvantage, compared with courses in other instructional periods, this laboratory course's student evaluation response rate was similar to all required courses in this year of the curriculum for all 3 years of the study. Usable student feedback might be obtained by prospectively using a questionnaire specifically designed for this course and independent of the automatic process employed by the college administration.

Efforts were made to ensure consistency among instructors in teaching and grading, but individual instructor differences are often difficult to control. One similarity noted among the instructors was the categories of grading that received the most G and U grades. Assessing a G or U required more effort on the part of the instructor than assigning an A, the default grade; therefore, it could be assumed that categories receiving more G and U grades are held in higher regard by the instructors than other categories. As such, it appears that the laboratory instructors felt that the following grading categories were the more important ones: pack preparation, aseptic technique, instrument handling, tissue handling, and knowledge of procedure. Alternately, it is possible that instructors found these categories easier to assess and, therefore, settled for the default grade in the other categories. The A was chosen as the default grade because faculty assumed that student performance should be adequate and did not want to risk an inadvertent U being given to someone with adequate performance. Ideally, a default of no score, forcing the instructor to enter a grade, should be used, but might not achieve the quick feedback that was intended. To further delineate the reason why some categories receive more Gs and Us than others, instructors could be asked to rank the grading categories in order of importance and in order of assessment difficulty. Categories determined to be of low importance could be eliminated from grading assessment, and improved rubric could make difficult categories easier to assess.

Instructors were asked to grade strictly during the first five laboratories to prepare students for the final two laboratories, which actually determined passage or failure. One could argue that a better tact would be to grade leniently at first and increase strictness as the students advanced in proficiency. However, grading students leniently could set the students up for failure if they were allowed to make mistakes in the preliminary laboratories that could result in failure in the last two laboratories. Grading strictly in the preliminary laboratories was aimed at building good surgical skills and developing good habits. Instead of grading leniently in the final two laboratories, the type of error that warranted failure was specifically defined as an error that would likely cause significant patient morbidity or damage to surgical instrumentation.

An attempt was made to make sure that residents did not outnumber faculty in laboratory assignments and that each student was assigned to each faculty surgeon by the end of the course. Since it sometimes became necessary to substitute surgery residents for faculty surgeons in the laboratory, the authors felt it important to compare faculty and resident grades to see if one group was more strict or lenient than the other. Apparently, both groups were comparable in strictness/leniency because there were no significant differences in the numbers of G and A grades. There were some differences between faculty and residents in the number of U grades; however, a lack of difference between faculty and residents was noted when G, A, and U grades were compared on a per-student basis.

Differences were more apparent when comparing individual instructors. Some instructors were more likely to give G or U grades, and few instructors gave a balanced number of G and U grades. Directions to grade strictly, particularly in Laboratories 1 through 5, did not result in an equivalent level of strictness among instructors. Therefore, the importance of careful scheduling to make sure that students are graded by a variety of instructors is reinforced. Additionally, the reason for the difference in level of strictness among instructors could be explored and addressed. For instance, if individual instructors are interpreting the grading criteria differently, more explicit written guidelines to define strictness could be developed.

Ways to decrease inter-grader disparities might include providing specified lists of what constitutes a U grade and making it easier for instructors to provide comments. Using the data accumulated to date, a list of what constitutes a U grade in each category could be developed. This list could be distributed to instructors along with an orientation regarding how to use it. To encourage comments, dropdown lists of corrective and complimentary comments could be developed based on what has been recorded to date. The dropdown lists of comments for common mistakes would facilitate time-efficient comment writing, which, in turn, would ameliorate tendencies for a grader to not give a U in order to avoid extra work.

Given the multitude of laboratory instructors necessary to teach this course and the subjective nature of performance assessment, establishing grading consistency poses a significant challenge. One could reorganize the laboratory structure to address grading consistency. For example, individual instructors could be assigned to the same students

for the duration of the course. Such would ensure consistency among an individual instructor's assigned students, but not among all students in the laboratory. Also, the advantage of exposing the students to a variety of instructor teaching styles would be lost.

There does not appear to be a preference for faculty over residents in terms of grading; however, this report did not evaluate quality of instruction. One might assume that instruction quality is better with faculty than with residents and thereby justify prioritizing the scheduling of faculty over residents in the laboratories. Scheduling based on anticipated quality of instruction requires further investigation because the opposite (i.e., better instruction by residents compared with faculty) could be true.

The number of G and U grades assigned could reflect the quality of the students as opposed to the strictness/leniency of the instructors. The number of G grades assigned in 2002 might suggest that the students in 2002 were overall more proficient in surgery than the students in the 2003 and 2004 laboratories, or the instructors could have graded more leniently in 2002 compared with 2003 and 2004. However, it should be noted that 2002 was the only year when U grades were given in one of the final two laboratories, necessitating a make-up laboratory for two students to avoid failing the course. Assigning U grades in the final two laboratories is more likely to occur with strict compared with lenient graders, but could occur with a lenient grader if the students were truly incompetent. It is difficult to make a strong conclusion about grading strictness or student proficiency based on only two U grades in the final laboratories. Looking at the progression of student grades from the first to last laboratory and comparing the number of G, A, and U grades in Laboratories 1, 5, and 7, it appears that each year student competency did increase as the laboratory course progressed. Laboratory 7 could be skewing this interpretation because instructors graded more strictly in Laboratories 1 and 5; however, Laboratory 5 grades did improve over those in Laboratory 1. In Laboratory 5 there were more Gs in 2002 and 2004 and less Us in 2003 compared with Laboratory 1.

The surgical faculty recognized the disadvantages of student assessment using a pass–fail system such as the S/U grading described here (Dietrick et al., 1991; Moss et al., 1978; Ravelli and Wolfson, 1999; Reznick et al., 1989), but designed this system because they felt it would be difficult to judge surgical performance in a letter grade fashion. The surgery faculty felt that they could subjectively conclude whether a student was average or above average or inadequate by observing surgical performance, but did not feel comfortable assigning numerical scores that would distinguish between A vs. B, B vs. C, and so forth. Certainly, the quizzes and final examination could be used to assign a letter grade, but the emphasis of this course is on the students' operative technique.

The S/U grading format available at the authors' university allows the recording of only S or U as the student's final grade. A student receiving a U in one of the final two laboratories received an unsatisfactory final grade, unless a G or A was earned in the make-up laboratory. A student receiving all G and A grades in the final two laboratories received a satisfactory final grade. The faculty included the G grade during the laboratories to acknowledge areas where the student was performing particularly well, thereby providing notes of encouragement. The positive reinforcement provided by the G grade would not be possible if the university's S/U grading format was directly applied during the laboratories.

This study was focused on instruction and grading provided by small animal surgery faculty and residents, but this laboratory also incorporates instruction in anesthetic techniques and ophthalmic surgery. Initially, anesthesiology and ophthalmology faculty declined to participate in grading, likely because of the smaller critical mass of faculty and residents in those disciplines compared with the number of available small animal surgery faculty and residents. However, in 2004, the ophthalmologists contributed quiz questions and the anesthesiologists indicated a desire to provide quiz questions in 2005.

This study has demonstrated effective use of a pass–fail grading mechanism in a didactic surgery laboratory. However, notable areas for improvement include more student feedback and increased consistency of grading among the varied instructors. A prospective student questionnaire with appropriate orientation and timing would enhance student feedback. More consistency among instructors in assigning grades might be achieved by further defining what level of good performance and what level of error warrants a G or U, respectively.

Conclusions

The novel grading mechanism described here was successfully employed during the first 3 years of its implementation. Efforts to improve the subjective grading component should include planned acquisition of student feedback and establishment of more consistency among graders**.** While objective criteria may not be enough to adequately assess overall performance in didactic surgery laboratories, consistency of subjective evaluation requires adherence to well-defined assessment criteria.

Acknowledgment

The authors would like to offer special thanks to Ms. Stephanie Bossaller, User Support Analyst Expert, and J.E. Creed, DVM, MS, Diplomate ACVS, Professor Emeritus, for their dedication and tireless efforts in employing the grading methods described in this manuscript.

References

Dietrick, J.A., M.T. Weaver, and H.W. Merrick. 1991. Pass/fail grading: A disadvantage for students applying for residency. Am. J. Surg. 162:63–66.

Greenfield, C.L., A.L. Johnson, M.W. Arends, and A.J. Wroblewski. 1993. Development of parenchymal abdominal organ models for use in teaching veterinary soft tissue surgery. Vet. Surg. 22:357–362.

Greenfield, C.L., A.L. Johnson, D.J. Schaeffer, and L.L. Hungerford. 1995. Comparison of surgical skills of veterinary students trained using models or live animals. J. Am. Vet. Med. Assoc. 206:1840–1845.

Griffon, D.J., P. Cronin, B. Kirby, and D.F. Cottrell. 2000. Evaluation of a hemostasis model for teaching ovariohysterectomy in veterinary surgery. Vet. Surg. 29:309–316.

Hedlund, C.S., G. Hosgood, and S. Naugler. 2002. Surgical education: attitudes toward animal use in teaching surgery at Louisiana State University. J. Vet. Med. Educ. 29:50–55.

Moss, T.J., E.C. Deland, and J.V. Maloney. 1978. Selection of medical students for graduate training: Pass/fail versus grades. N. Engl. J. Med. 299:25–27.

Olsen, D., M.S. Bauer, H.B. Seim, and M.D. Salman. 1996. Evaluation of a hemostasis model for teaching basic surgical skills. Vet. Surg. 25:49–58.

Ravelli, C., and P. Wolfson. 1999. What is the "ideal" grading system for the junior surgery clerkship? Am. J. Surg. 177:140–144.

Reznick, R.K., J.A. Colliver, R.G. Williams, and J.R. Folse. 1989. Reliability of different grading systems used in evaluating surgical students. Am. J. Surg. 157:346–349.

Smeak, D.D., L.N. Hill, M.L. Beck, C.A. Shaffer, and S.J. Birchard. 1994. Evaluation of an autotutorial-simulator program for instruction of hollow organ closure. Vet. Surg. 23:519–528.

About the author...

Dr. Mann earned an AAS in Veterinary Technology at Morehead State University in 1977 and DVM from the Ohio State University in 1982. He completed a small animal rotating internship at the University of Missouri–Columbia in 1983 and small animal surgical residency and M.S. at Texas A&M University in 1986. Dr. Mann served as an assistant professor at Auburn University from 1986 to 1988, and joined the University of Missouri–Columbia in 1988 where he rose to full professor in 2006. Dr. Mann is board-certified in both the American College of Veterinary Surgeons and the American College of Veterinary Emergency and Critical Care.

F. A. (Tony) Mann

articles

APPENDIX

For the duration of the authors' employment at the college of veterinary medicine, training veterinary students in the art and science of small animal surgery in a didactic setting has been a high priority. Successful completion of a didactic surgical laboratory experience utilizing live animals and cadavers has remained a prerequisite for the third- and fourth-year clinical rotations. Before 1992 there was one didactic small animal surgery course (2.5 credit hours) taught during an 8-week instructional period every May–June. This course consisted of lectures in orthopedics, soft tissue surgery, and neurosurgery. Also, a didactic surgical laboratory was held three afternoons per week (each student having 1 day of laboratory per week) using live dogs for soft tissue surgical procedures and cadavers for orthopedic procedures. Euthanasia was performed at the end of soft tissue procedures and the cadavers were held in cold storage until the following week for the orthopedic procedures such that each week alternated between terminal live-animal surgery and cadaveric surgery. Typically, ophthalmic surgical procedures were performed on cadavers at the end of the shortest orthopedic laboratory.

In 1992, two additional credit hours were assigned to small animal surgery. These credit hours were assigned to an 8-week instructional period in March–April during which basic surgical principles and soft tissue surgery lecture topics were taught. The 2.5 credit hours in May–June were retained, but all lecture material in May–June was orthopedic, and the didactic surgery laboratory continued as described above as part of the orthopedic lecture course. Before 1992, students received one letter grade for small animal surgery (lecture and laboratory), but this letter grade was based on performance on examinations that covered the lecture material. Other than pre-laboratory quizzes worth extra points, the laboratory had no effect on the final letter grade, albeit students were required to attend the laboratory and make a conscientious effort in the laboratory to pass the small animal surgery course. One year there was an attempt to assign laboratory grades to the students for their performance in the laboratories, but inability for laboratory instructors to cover all assigned tables with consistency resulted in inequity of assigned grades. Therefore, the laboratory portion of the May–June surgery course reverted to a pass–fail status. The pass–fail status of the laboratory was of little consequence to the required letter grade because the letter grade was earned through testing of lecture material. Nonetheless, the surgical faculty recognized a need to more critically evaluate student performance in the laboratory.

The practice of assigning a letter grade based on the testing of lecture material while the laboratory was considered pass–fail continued beyond 1992, but the letter grade was achieved through testing of orthopedic lecture material. In 1999, the surgical faculty began discussions regarding the separation of the laboratory from the orthopedic lectures to provide two separate courses. These discussions were catalyzed by the fact that the faculty member in charge of coordinating the laboratory was not the same person who was the lead instructor for the overall course. The laboratory course required equal, if not more, time for preparation, organization, and implementation compared with the orthopedic lecture portion. Further, the surgical faculty felt that the didactic surgery laboratory was important enough to be conducted independently from the orthopedic lectures and should be evaluated separately. However, the major impediment for moving forward with this split was the challenge of how to effectively grade the students. The surgical faculty felt that written tests could play a role, and could certainly segregate students within a letter grading scale, but the emphasis of the laboratory should be in the evaluation of surgical etiquette, aseptic technique, competence in performing the assigned procedures, and other intangibles for which it would be difficult to objectively assign a letter grade.

In 2001, the didactic surgical laboratory was split from the orthopedic lecture course and maintained in the May–June instructional period with 0.5 credit hour assigned to the surgical laboratory and 2.0 credit hours assigned to the orthopedic lecture course. The surgical faculty redoubled their efforts to devise a grading system for the laboratory.

The didactic surgery laboratory course was offered every May–June from 2001 to 2004 in an 8-week instructional period. Satisfactory/unsatisfactory grading was used in 2001, but the refined method of doing so reported here was not implemented until 2002 through 2004.