A Change to Pass/Fail Grading in the First Two Years at One Medical School Results in Improved Psychological Well-Being

Robert A. Bloodgood, PhD, Jerry G. Short, PhD, John M. Jackson, MS, and James R. Martindale, PhD

Abstract

Purpose
To measure the impact of a change in grading system in the first two years of medical school, from graded (A, B, C, D, F) to pass/fail, on medical students' academic performance, attendance, residency match, satisfaction, and psychological well-being.

Method
For both the graded and pass/fail classes, objective data were collected on academic performance in the first- and second-year courses, the clerkships, United States Medical Licensing Examination (USMLE) Steps 1 and 2 Clinical Knowledge (CK), and residency placement. Self-report data were collected using a Web survey (which included the Dupuy General Well-Being Schedule) administered each of the first four semesters of medical school. The study was conducted from 2002 to 2007 at the University of Virginia School of Medicine.

Results
The pass/fail class exhibited a significant increase in well-being during each of the first three semesters of medical school relative to the graded class, greater satisfaction with the quality of their medical education during the first four semesters of medical school, and greater satisfaction with their personal lives during the first three semesters of medical school. The graded and pass/fail classes showed no significant differences in performance in first- and second-year courses, grades in clerkships, scores on USMLE Step 1 and Step 2CK, success in residency placement, and attendance at academic activities.

Conclusions
A change in grading from letter grades to pass/fail in the first two years of medical school conferred distinct advantages to medical students, in terms of improved psychological well-being and satisfaction, without any reduction in performance in courses or clerkships, USMLE test scores, success in residency placement, or level of attendance.

Acad Med. 2009; 84:655–662.

Editor's Note: Commentaries on this article appear on pages 545 and 548.

In recent years, U.S. medical schools have used a variety of grading systems, particularly in the first two years of medical school. The most recent data available from the Association of American Medical Colleges Web site (http://services.

Dr. Bloodgood is professor of cell biology, University of Virginia School of Medicine, Charlottesville, Virginia.

Dr. Short is associate dean for medical education support and professor of medical education, University of Virginia School of Medicine, Charlottesville, Virginia, and professor of education, Curry School of Education, University of Virginia, Charlottesville, Virginia.

Mr. Jackson is director of educational technology and assistant professor of medical education, University of Virginia School of Medicine, Charlottesville, Virginia.

Dr. Martindale is assistant professor of medical education, University of Virginia School of Medicine, Charlottesville, Virginia.

Correspondence should be addressed to Dr. Bloodgood, Department of Cell Biology, University of Virginia School of Medicine, P.O. Box 800732, Charlottesville, VA 22908-0732; telephone: (434) 924-1739; fax: (434) 982-3912; e-mail: (rab4m@virginia.edu).

aamc.org/currdir/section1/grading1.cfm) show the following breakdown of the grading systems used by medical schools for the required basic sciences portion of the curriculum: two intervals (usually pass/fail) = 40 schools, three intervals (usually pass/fail/honors) = 35 schools, four intervals (usually pass/fail/honors/high honors) = 32 schools, and five intervals (usually A, B, C, D, F) = 26 schools. These data suggest that there is currently no consensus on the most appropriate grading system in the early years of medical school. In addition, there are few useful data available to guide decision making in this area.[1–5]

Medical school is inherently stressful.[6–8] The principal attraction in moving toward a pass/fail grading system lies in the expectation that it will improve students' psychological well-being (reduce stress and anxiety), decrease competitiveness, and promote cooperative learning. However, for medical schools contemplating a change from a traditional graded (A, B, C, D, F) system to a pass/fail system, concerns may include (1) a decline in attendance at scheduled educational activities, (2) a decline in academic performance, (3) a decline in United States Medical Licensing Examination (USMLE) Step 1 scores, and (4) reduced success in residency placement.

A University of Virginia School of Medicine faculty committee examined the literature on medical student grading systems (and its relationship to academic performance), medical student well-being, and residency placement. The committee surveyed current medical students and faculty regarding their preferences for grading systems and gathered data on grading policies at other medical schools. Based on these preliminary measures, in the spring of 2002, the decision was made to switch from a five-interval (A, B, C, D, F) to a two-interval (pass/fail) grading system for each individual course in the first two years of medical

Supplement digital content is available for this article. (http://links.lww.com/A1172)

school. The change in grading system was introduced with the class of 2007 (the class entering medical school in the fall of 2003).

The faculty committee that recommended the switch to pass/fail grading hypothesized that the change from a graded to a pass/fail system would result in no change in attendance at scheduled academic activities, no change in academic performance in courses and clerkships, no change in performance on USMLE Step 1 and Step 2 Clinical Knowledge (CK) exams, and no change in residency placement success. However, it was predicted that this change in grading system would improve students' well-being, increase satisfaction with their medical education, and affect their time utilization (e.g., allow more "risk taking" in terms of time allocation to activities not directly related to the medical curriculum). A comprehensive assessment of the impact of the change in the grading system at the University of Virginia School of Medicine has been conducted, the results of which we report here. R.A.B. cochaired the grading study committee, designed the study, collected the data, and wrote the paper. J.G.S. was a member of the grading study committee and assisted in study design and data analysis. J.M.J. designed the online survey tool and assisted in data collection and processing. J.R.M. did all of the statistical analysis of the data.

Method

Our protocol involved the comparison of two medical school classes at the University of Virginia School of Medicine: the class of 2006 (n = 141), the last class under the five-interval grading system in the first two years of medical school, and the class of 2007 (n = 140), the first class under the pass/fail grading system in the first two years of medical school. The medical curriculum experienced by these two classes (to be referred to as the graded and pass/fail classes) was essentially identical; both classes took the same array of courses (eight courses in year 1 and six courses in year 2). Both classes experienced a five-interval grading system in the clerkships and a pass/fail grading system in the fourth-year electives portion of the curriculum. Grades on individual clerkships reflect a combination of (1) evaluations by clinical

preceptors (faculty and residents) and (2) the score on the National Board of Medical Examiners "shelf exam." The clerkships do not use a fixed distribution for the assignment of letter grades. No change in clerkship grading procedures occurred between the graded and pass/fail classes.

While the class of 2007 studied under a pass/fail grading system at the level of each individual course, cumulative honors were awarded to 20% of the medical class at the end of the second year; honors were calculated by a simple average of all percentage scores in all courses. The committee that recommended a change to pass/fail grading did not include any recommendation for cumulative honors; this was added by the dean of the school of medicine as a compromise to gain the support of a small minority of department chairs who wanted a means to recognize superior performance (presumably for the purpose of residency placement). The course directors maintained percentage scores for course performance for each student (a blend of exam scores, quiz scores, and lab performance scores that varied for each course) even when only "P" or "F" grades were awarded to students.

We collected students' course performances, USMLE Step 1 scores, clerkship grades, USMLE Step 2CK scores, and residency program quality measures from existing databases. Self-report data were collected by means of a Web-based survey (survey available as supplemental Appendix online, (http://links.lww.com/A1172), which we administered to the two classes once each semester for the first four semesters of medical school; the surveys were administered to all 141 members of the graded class and to all 140 members of the pass/fail class. Each class received the surveys at the same time in the semester, and each administration fell at least two weeks away from any vacation period or exam period; each survey was available on the Web for one week. The survey was completely deidentified, and there were no incentives provided to medical students for completing the surveys. The response rate on these voluntary surveys varied with administration from 44% (n = 62) to 75% (n = 103); the subset of students completing the survey varied

with administration. The survey contained the complete Dupuy Schedule of General Well-Being[9] as well as additional questions on satisfaction with medical school, satisfaction with personal life, attendance at scheduled academic activities, and time utilization. The Dupuy Schedule, developed by Harold J. Dupuy, has been widely used in clinical studies and has been well validated,[9–11] although we are not aware of its previously having been used in undergraduate medical education. It uses 18 questions that gather data on anxiety, depression, positive well-being, self-control, vitality, and general health.[9] As part of a larger Web-based survey, this instrument was administered to both the graded and pass/fail classes in each of the first four semesters of medical school. It was administered at approximately the same time in each semester for each class, and the administration of the survey fell at least two weeks away from any vacation period or exam period. The Likert questions on the Dupuy Schedule were assigned numeric scores, and a cumulative Dupuy Schedule score (maximum of 124 points) was calculated for each student for each administration of the survey. Dupuy Schedule subscores were calculated as indicated in Exhibit 5.12 in Dupuy et al.[9] IRB approval for this study was obtained before the commencement of any data collection and was renewed annually.

Statistical analyses

Initial demographics for both classes were calculated and included number of students in each class, mean age, mean undergraduate grade point average, mean MCAT scores for the Biology, Physics, and Verbal Reasoning sections, gender, minority status, proportion of students who were University of Virginia undergraduates, and the percentage who were Virginia residents. Two-sample independent *t* tests were used to compare the two classes on mean MCAT Biology, Physics, and Verbal Reasoning test components. Chi-square tests of independence were used to look at the association between class and gender, minority status, percentage of class, attendance at the University of Virginia for undergraduate degree, and Virginia residency status.

Mean overall course grades were analyzed by class and by gender within class.

Means and standard deviations were calculated for each of the three group comparisons, and a two-sample independent t test was used to look for statistically significant mean differences between the overall graded and pass/fail class performances, as well as between the graded and pass/fail class performances for males only and for females only.

Overall measures of medical student well-being obtained from the Dupuy Schedule were compared for the graded and pass/fail classes. Means and standard deviations were computed, and mean group comparisons were addressed for statistical significance by the use of two-group independent t tests. Additionally, overall mean group comparisons were made for each class, broken down by gender. Comparisons were made for each of the first four semesters in medical school. The Dupuy Schedule was also broken down into its individual subscale components (anxiety, depression, positive well-being, self-control, vitality, and general health), and mean group comparisons were made for each subscale for the graded versus pass/fail classes for each of the first four semesters of medical school. Means and standard deviations were computed, and mean group comparisons were made using two-group independent t tests.

Mean board certification exam pass rates were compared for the residency programs to which the graded and pass/fail classes matched. Five-year average pass rates for individual residency

training programs in internal medicine, family medicine, pediatrics, and general surgery were obtained from the appropriate academy Web sites. These were the only residency programs for which board certification exam rates are available at the level of individual training programs, and hence our conclusions about residency placement success must be qualified as applying only to these four types of residency programs (representing about half of our medical graduates). Means and standard deviations were calculated, and a two-group independent t test was used to look for mean group differences.

Cohen effect sizes as well as 95% confidence intervals were calculated for all mean differences. An a priori alpha level of .05 was used for all inferential tests. To control for inflated type I error rates due to multiple significance tests, a Bonferroni correction procedure was applied to each set of significance tests. All statistical analyses were performed using SPSS version 13 (SPSS Inc., Chicago, Illinois).

Results

Comparison of the characteristics of the graded and pass/fail medical classes

The characteristics of the two University of Virginia medical school classes (class of 2006 and class of 2007) used in this study are shown in Table 1. They were well matched with the exception of gender composition; the graded class had

38% male and 62% female students, and the pass/fail class had 54% male and 46% female students, a statistically significant difference ($P = .024$). Because of this gender disparity between the graded and pass/fail classes, all across-class comparisons were performed three ways: male and female combined, only male, and only female.

Academic performance

Performance in medical school courses for the first two years of the curriculum is shown in Table 2. Even though the class of 2007 was under a pass/fail grading system in each course, course directors calculated percentage scores for each student for each course on the basis of performance in exams, quizzes, laboratories, etc. Students were aware of these percentage scores. For each student in each of the two classes, we averaged all of the individual course scores for the first two years of medical school to generate an individual mean; from these individual student means, we generated a class mean. There was no statistically significant difference between the two classes in terms of academic performance on all first- and second-year courses combined, even when the data were broken down by gender (Table 2). When course performance in first- and second-year courses was analyzed separately, there were still no significant differences in academic performance between the graded and pass/fail classes ($P = .651$ for all first-year courses and $P = .397$ for all second-year courses).

Performance in clinical clerkships did not differ between the graded and pass/fail classes. For both classes, the clinical clerkships used letter grades, including plus and minus; two different schemes for converting letter grades to numeric scores were compared. One scheme is used by this medical school for the selection of medical students for membership in Alpha Omega Alpha, the national medical honor society, and is based on a 13-point scale (A+ = 13, A = 12, A− = 11, etc.). The mean clerkship grade was 10.59 for the graded class and 10.69 for the pass/fail class ($P = .165$). The other scheme is a four-point scale in which an A (A+, A, A−) is worth four points, a B (B+, B, B−) is worth three points, etc. The mean clerkship grade was 3.54 for the graded class and 3.57 for the pass/fail class ($P = .426$).

Table 1

Demographic Characteristics of Two Medical School Classes, From a Study of Graded Versus Pass/Fail Grading Systems, University of Virginia Medical School, 2007

Characteristic	Class of 2006: Five intervals (A, B, C, D, F)	Class of 2007: Two internals (pass/fail)	P value
Students in the class (no.)	141	140	
Age (mean years/months)	22/4	22/4	
Undergraduate GPA (points)	3.66	3.70	.150
MCAT score—biology (points)	10.73	10.83	.549
MCAT score—physics (points)	10.75	10.94	.267
MCAT score—verbal reasoning (points)	10.22	10.31	.587
Male (%)	38	54	.024
Female (%)	62	46	.024
Underrepresented minorities (%)	5.1	5.7	.770
University of Virginia undergrads (%)	26.1	25.7	.990
Virginia residents (%)	62.3	65.7	.570

Table 2

Student Academic Performance Data by Gender, From Study of Graded Versus Pass/Fail Grading Systems, University of Virginia Medical School, 2007

Students	No.	Mean (SD)	95% CI	ES*	P value
All			−1.38 to 0.744	.071	.558
Graded	137	87.14 (4.46)			
Pass/fail	138	87.46 (4.47)			
Male			−1.72 to 1.43	.032	.959
Graded	53	88.06 (4.82)			
Pass/fail	74	88.02 (4.19)			
Female			−1.59 to 1.33	.028	.738
Graded	84	86.56 (4.13)			
Pass/fail	64	86.81 (4.72)			

* ES indicates Cohen effect size.

USMLE scores

In terms of first-time performance on USMLE Step 1 (a measure of academic performance in the first two years of the curriculum), the pass/fail class performed as well as the graded class. The graded class had a mean of 222 ($\pm$20), and the pass/fail class had a mean of 226 ($\pm$21), out of a possible 260. The comparable national means for those two years were 216 ($\pm$24) and 218 ($\pm$23). A two-tailed t test showed no statistically significant difference ($P = .101$) between the two classes. On the USMLE Step 2CK exam, the pass/fail class again performed as well as the graded class. The graded class had a mean of 231 ($\pm$21), and the pass/fail class had a mean of 236 ($\pm$21), out of a possible 260. The comparable national means for those two years were 221 ($\pm$24) and 225 ($\pm$24). A two-tailed t test showed no statistically significant difference ($P = .060$) between the two classes.

Residency placement success

Given the importance of the residency match, the concern that a change in medical school grading in the first two years of medical school from a graded to a pass/fail system might have a negative impact on residency placement, and the absence of any useful measure (other than percent of total medical students in a class who match to any residency program) for accessing the overall quality of the residency programs to which a group of students matches, we examined a number of possible measures (including various medical school rankings). The most relevant and objective measure that we could identify for the quality of individual residency programs was based on the board certification pass rates (five-year averages) for the residents in the specific residency programs to which the medical students in this study matched. These scores are available for all residency programs in internal medicine, family medicine, pediatrics, and general surgery on the respective board Web sites. For these four categories of residency programs combined, there was no significant difference in the mean board certification pass rates for the residency

Table 3

Comparison of Medical Student Well-Being Between Graded and Pass/Fail Medical School Classes, From a Study of Graded Versus Pass/Fail Grading Systems, University of Virginia Medical School, 2007

Students	Semester	Graded class			Pass/fail class			95% CI	ES[†]	P value
		No.	Dupuy Schedule score*	SD	No.	Dupuy Schedule score*	SD			
All										
	1	81	80.56	16.40	103	91.46	14.46	−15.37 to −6.43	.122	<.001
	2	106	85.74	15.29	96	91.97	17.73	−10.78 to −1.68	.376	.008
	3	62	77.56	14.79	101	87.13	16.64	−14.62 to −4.52	.607	<.001
	4	88	77.74	15.13	100	79.02	18.89	−6.22 to 3.66	.074	.610
Female										
	1	47	77.51	14.67	48	91.44	14.28	−19.75 to −8.11	.962	<.001
	2	63	82.37	14.43	48	90.52	19.47	−14.45 to −1.85	.475	.013
	3	36	75.25	14.61	48	84.58	14.77	−15.68 to −2.98	.635	.005
	4	50	76.42	15.16	48	74.73	19.11	−5.13 to 8.51	.098	.628
Male										
	1	34	84.79	17.91	55	91.47	14.75	−13.53 to 0.17	.407	.059
	2	43	90.67	15.32	48	93.42	15.87	−9.18 to 3.68	.176	.405
	3	26	80.77	14.72	53	89.43	17.99	−16.64 to −0.68	.526	.037
	4	38	79.47	15.12	52	82.98	17.97	−10.55 to 3.53	.211	.332

* The Dupuy Schedule of General Well-Being is a 124-point scale that measures anxiety, depression, positive well-being, self-control, vitality, and general health. In all cases, a larger number denotes improved well-being.
[†] ES indicates Cohen effect size.

programs to which the medical students in the graded and pass/fail classes were matched (Table 5).

Psychological well-being

Table 3 shows the mean Dupuy Schedule scores for each class for each of the four administrations. The pass/fail class exhibited a statistically significant increase in well-being ($P < .01$) compared with the graded class for the first three semesters of medical school. Although the pass/fail class still exhibited a higher level of well-being than the graded class in the fourth semester of medical school, the difference was not statistically significant. When the data were broken down by gender, the same pattern of greater well-being in the pass/fail class relative to the graded class for the first three semesters of medical school was observed for male and female medical students. This effect was statistically significant for females in the first three semesters of medical school ($P < .05$) but statistically significant for males only in the third semester ($P < .05$).

Because the Dupuy Schedule assigns questions to six different categories (anxiety, depression, positive well-being, self-control, vitality, and general heath), we calculated subscores for these six criteria for each student in each class for each administration of the survey (Table 4). For the first three semesters of medical school, the pass/fail class exhibited higher scores on all six subscores relative to the graded class. For all subscores, higher scores represented a better state of well-being (i.e., less anxiety, less depression, more positive well-being, more self-control, more vitality, and better general health). Table 4 shows that, in semester 1, the improvement in well-being status for the pass/fail class was statistically significant for all six subscore categories. For semesters 2 and 3, the improvement in well-being for the pass/fail class was statistically significant for four of the six categories (anxiety, depression, positive well-being, and vitality). For semester 4, there were no statistically significant differences between the two classes for any of the subscores.

Cumulative honors and stress

There was one somewhat unusual feature of the grading system experienced by the pass/fail class: the awarding of cumulative honors in the basic sciences at the end of the first four semesters of medical school. Although every individual course used a straight pass/fail grading system, individual student performances in the entire array of first- and second-year courses were averaged, and the top 20% of the class received cumulative honors. Anecdotal discussion with medical students under the pass/fail grading system suggested that many students chose a strategy early in medical school either to work actively toward achieving the cumulative honors or to ignore the cumulative honors component of the pass/fail grading system. This stimulated a separate survey (dealing only with cumulative honors) of the pass/fail class, which was administered after completion of the first two years of medical school. Of those medical students who felt they had made a conscious decision to seek cumulative honors, 70% felt that this decision resulted in greater stress, 30% felt this decision had no effect on level of stress, and 0% felt that this decision resulted in lower stress during the first four semesters of medical school. Of those who felt that they had made a conscious decision to not seek cumulative honors, 0% felt that this decision resulted in greater stress, 8% felt that this decision had no effect on perceived stress, and 92% felt that this decision resulted in lower stress.

Satisfaction, attendance, and time utilization

The major survey instrument used in this study, besides the Dupuy Schedule, also gathered self-reported data from the medical students on satisfaction with medical school, satisfaction with personal life, attendance at scheduled academic events, and time utilization. The pass/fail class exhibited a statistically significant increase ($P < .05$) in "satisfaction with the quality of my medical education" for each of the first four semesters of medical school relative to the graded class, although this effect was more robust in the first year of medical school than in the second year. The pass/fail class also exhibited a statistically significant increase ($P < .05$) in "current satisfaction with my personal life during the last month" for the first three semesters of medical school relative to the graded class; this effect disappeared in the fourth semester of medical school, as did the increased well-being (based on Dupuy Schedule) shown in Tables 3 and 4. There

Table 4

Comparison of Dupuy Schedule Subscores Between Graded and Pass/Fail (P/F) Classes, From a Study of Graded Versus Pass/Fail Grading Systems, University of Virginia Medical School, 2007*

	Semester 1			Semester 2			Semester 3			Semester 4		
Subscore	Graded (n = 82)	P/F (n = 103)	P value	Graded (n =107)	P/F (n = 96)	P value	Graded (n = 62)	P/F (n = 102)	P value	Graded (n = 88)	P/F (n = 100)	P value
Anxiety	15.98	18.14	.002[‡]	17.65	19.01	.048[†]	14.55	17.02	.001[‡]	14.20	14.08	.858
Depression	15.89	17.62	<.001[§]	16.65	17.61	.049[†]	15.08	16.92	.001[‡]	15.35	15.56	.705
Positive well-being	11.02	13.02	<.001[§]	12.20	13.09	.030[†]	10.74	12.37	<.001[§]	10.40	10.59	.674
Self-control	14.12	15.51	<.001[§]	14.45	15.10	.126	14.40	15.13	.078	14.42	14.61	.632
Vitality	12.15	14.60	<.001[§]	13.31	15.16	.001[‡]	11.95	14.10	<.001[§]	12.06	12.88	.111
General health	11.48	12.56	.016[†]	11.24	11.99	.152	10.84	11.25	.472	11.31	11.30	.989

* The Dupuy Schedule of General Well-Being is a 124-point scale that measures anxiety, depression, positive well-being, self-control, vitality, and general health.
[†] $P < .05$.
[‡] $P < .01$.
[§] $P < .001$.

Table 5

Residency Placement Success Judged by the Board Certification Pass Rates for Individual Residency Programs to Which the Medical Students Matched, From a Study of Graded Versus Pass/Fail Grading Systems, University of Virginia Medical School, 2007

	95% CI	Graded class (n = 63)	Pass/fail class (n = 52)	ES[†]	P value
Board exam pass rates for residency programs (%)*	−3.87 to 3.57			.014	.937
Mean (SD)		90.34 ± 10.66	90.49 ± 9.45		
Range		59–100	59–100		

* Programs were in internal medicine, family medicine, pediatrics, and general surgery.
† ES indicates Cohen effect size.

were no statistically significant changes in self-reported attendance at scheduled academic events or in self-reported time utilization (voluntary clinical activities, independent scholarly activities, community service activities, exercise-related activities, and leisure activities other than exercise) between the two classes during the first four semesters of medical school.

Discussion

The literature on medical education contains very few useful data that address the effects of the grading system in the early years of medical school on factors such as course performance, USMLE scores, time utilization, attendance, residency placement success, psychological well-being, satisfaction with medical education, and competitiveness among students. This leaves medical schools with little basis for making informed decisions about choosing or changing their grading systems.

Our study has demonstrated that a change from a five-interval (A, B, C, D, F) to a two-interval (pass/fail) grading system in the first two years of medical school at the University of Virginia School of Medicine was not associated with a decline in students' academic performance (course performance in the first two years of the curriculum, USMLE Step 1 scores, clerkship grades, or USMLE Step 2 CK scores). This is consistent with previous data from the University of Michigan Medical School showing that a change from a four-interval (honors, high pass, pass, fail) to a two-interval (pass/fail) grading system in the first year of medical school resulted

in no statistically significant change in performance in the first-year gross anatomy course.[2] This was also the case for the gross anatomy course in the present study. On the other hand, when the Mayo Medical School switched from a five-interval (A, B, C, D, F) to a three-interval (pass, marginal pass, fail) grading system in the first year of medical school only, a statistically significant decrease in performance on written exams occurred in the first-year gross anatomy course.[4] Robins et al,[2] when comparing medical student performance in basic science courses at the University of Michigan under a four-interval (honors, high pass, pass, fail) versus a two-interval (pass/fail) grading system, found that medical students' performance on their last course examination correlated with the average of all previous examinations in that particular course, suggesting that the students did not "slack off" at the end of a course, even when already assured a passing grade based on their performance on the previous exams. Supporting this, we found no change in students' self-reported lecture attendance between the graded and pass/fail classes for any of the first four semesters of medical school. Rohe et al[5] found no statistically significant change in USMLE Step 1 scores when comparing a five-interval and a three-interval grading system at the Mayo Medical School. When looking at the credentials of applicants to the Northwestern University internal medicine residency program, Hughes et al[12] found that average USMLE Step 1 scores for students from schools with a two-interval (pass/fail) grading system exceeded those of students from schools with either five-interval (A, B, C, D, F) or

three-interval (high pass, pass, fail) grading systems.

The data from the current study indicate that the major benefit to be derived from a change from a five-interval (A, B, C, D, F) to a two-interval (pass/fail) grading system in the first two years of the medical school curriculum lies in the area of psychological well-being. Using the Dupuy Schedule,[9] which has questions that address anxiety, depression, positive well-being, self-control, vitality, and general health, we observed that the change to a pass/fail grading system resulted in a statistically significant improvement in students' well-being (mean cumulative Dupuy Schedule score) for each of the first three semesters of medical school (Table 3). When the data were broken down by gender, we again saw the same pattern of increased well-being for each of the first three semesters of medical school for both males and females. For females, this increase was statistically significant for semesters 1 through 3; for males, this increase was statistically significant only for semester 3. It is interesting to note that the switch to the pass/fail system was associated with a greater increase in well-being among females than among males in the first year (first two semesters). When the Dupuy Schedule well-being data were broken down by subscore, the change to pass/fail resulted in an increase in well-being for all subscores for each of the first three semesters (Table 4). For semester 1, these increases in well-being were significant for all subscores; for semesters 2 and 3, the increases were significant for anxiety, depression, positive well-being, and vitality. Just as the cumulative Dupuy scores showed no significant change (with change in grading system) in the fourth semester, the same applied to all six of the subscores.

The pass/fail grading system implemented by the University of Virginia School of Medicine effective with the class of 2007 had one unusual component: cumulative honors awarded to 20% of the class based on overall academic performance in the first two years of the curriculum. Whereas Dupuy Schedule data showed a significant decrease in stress associated with the switch to pass/fail grading, a separate survey of the class of 2007 on the impact of cumulative honors suggests that those

students who consciously chose to seek cumulative honors experienced greater stress relative to those students who consciously choose to not pursue cumulative honors. These data suggest that the greater well-being associated with a change to pass/fail grading that we have documented for the first three semesters of medical school might have been even greater had there not been the cumulative honors component of the pass/fail grading system. After two classes under the pass/fail grading system, we dropped the cumulative honors component, with the support of the medical students.

The greater well-being associated with a switch to pass/fail grading documented in the present study is consistent with other recent data. Rohe et al[5] studied the effect of a change in grading system from a five-interval (A, B, C, D, F) to a three-interval (pass/marginal pass/fail) system in the first year only at the Mayo Medical School. At the end of the first year, the students on the three-interval grading system exhibited less stress, better overall mood, and greater group cohesion; these levels of stress and group cohesion continued through the second year of the curriculum, even though a five-interval grading system was used. Rohe et al[5] found no correlation between test anxiety and the grading system in their study, even though examinations are considered by medical students to be among the greatest sources of stress in medical school.[8,13,14] Robins et al[2] cited student survey responses suggesting that a switch from a four-interval (honors, high pass, pass, fail) to a two-interval (pass/fail) grading system at the University of Michigan eased anxiety and reduced competition while encouraging student cooperation. Numerous comments on the open-ended question in the survey we used indicated that students in the pass/fail class felt that pass/fail grading reduced competition among medical students as well as stress on individual medical students.

It is of interest to consider why the benefit of greater well-being associated with a change from a graded to a pass/fail system disappeared in the fourth semester of medical school. It should be noted that the fourth semester of medical school showed the lowest level of well-being for the pass/fail class, a level comparable with both the third and fourth semesters of medical school for the graded class. The fourth semester showed declines in students' self-reported lecture attendance for both the pass/fail and graded classes (data not shown). Survey data from our medical students suggest that the impending USMLE Step 1 exam, coupled with cumulative course examinations, was responsible for both the decline in well-being (increased stress) and the decreased lecture attendance, since decreased lecture attendance freed up time to study for USMLE Step 1 and the cumulative course exams, which are used primarily at the end of the second year of the curriculum. This is consistent with the literature: Kidson and Hornblow[13] reported that the examinations marking the end of preclinical training at Monash University in Australia were rated by the majority of medical students as having provoked extremely high anxiety. Rosenthal et al[8] found that entry into medical school and the point just before the end of the second-year exams were the most stressful times during the first 30 months at the University of Tennessee College of Medicine. Clark and Zeldow[15] showed that depressive symptoms in medical students were highest at the end of the second year of medical school.

For each of the first four semesters of medical school, medical students in the pass/fail class rated their current level of satisfaction with the quality of their medical education higher than did the students in the graded class. This is consistent with observations of Robins et al,[2] who found that medical students at the University of Michigan Medical School on a two-interval (pass/fail) grading system in the first year of medical school were more satisfied with their evaluation and examination system and the learning environment than were students on a three-interval (honors, pass, fail) grading system. We found that the switch to a pass/fail grading system was associated with greater satisfaction with personal life for the first three semesters of medical school.

One of the areas of major concern when considering any change in grading system in a medical school is the possible impact on residency placement success. As far as we are aware, the only measures of success in the residency match that have been used by medical schools, by the National Residency Match Program, and in other research studies are the percentage of the total class that matched to any residency program, the percentage of students who matched to their first choice, or the percentage of students who matched to one of their first three choices. The present study is the first to quantitate residency placement "success" for a medical class using an objective measure (board certification pass rates) to assess the quality of the residency programs to which the medical students in the class were matched. Using this measure, there was no statistically significant difference in residency placement success between the graded and the pass/fail classes. There is considerable literature that suggests that grades in the first two years of medical school are not given much weight in the residency selection process.[16–19] A somewhat separate issue is whether medical students from medical schools with pass/fail grading systems suffer any competitive disadvantage in residency selection; the literature in this area is somewhat confusing because of differences in terms of which portion of medical school was pass/fail in the different studies (often not indicated). Tardiff[20] found that 73% of residency directors did not give preference to students from either graded or pass/fail schools. However, there were differences across specialties in the percentage of residency directors who gave preference to students from graded schools, ranging from a low of 6% in psychiatry programs to a high of 42% in surgery programs. Hughes et al[12] studied the selection of internal medicine residents at Northwestern University and found that "the overall impact of the grading system as a factor in the selection of residents seems to be minimal."

Although it is not a formal part of this study, we feel that we would be remiss not to comment on the potential impact of a change to pass/fail grading on medical admissions. In the years just before the change to pass/fail grading, surveys of students who rejected offers of admission to the University of Virginia School of Medicine often cited the grading system as a factor in their decision. In a prematriculation survey of the medical class entering in the fall of 2007 (under a pass/fail system, after we had eliminated cumulative honors), 81% said that the grading system was somewhat to very important in their

decision to accept the offer of admission from the University of Virginia School of Medicine.

In summary, our findings have shown that a change from a five-interval (A, B, C, D, F) to a two-interval (pass/fail) grading system at the University of Virginia School of Medicine did not result in any decline in academic performance (on first- and second-year courses, clerkships, and USMLE Step 1 and Step 2CK), attendance at scheduled academic activities, or residency placement success (as measured by the quality of the residency programs to which students were matched). This change to a pass/fail grading system was accompanied by a statistically significant improvement in psychological factors related to anxiety, depression, positive well-being, self-control, vitality, and general health in the first three semesters of medical school. These data on performance and psychological well-being are nicely reflected in one student survey comment that the pass/fail system "directly influences my stress level without compromising my effort level." The students in the pass/fail class reported increased satisfaction with the quality of their medical education in their first four semesters of medical school and with their personal life in their first three semesters of medical school.

Possible limitations of the current study include the facts that (1) survey response rates were lower for the graded class than the pass/fail class for three of the four semesters assessed, (2) the conclusion of similar success in the residency match for the graded and pass/fail classes is based on a portion of the study sample because data are available only for certain types of residency programs, (3) the graded and pass/fail classes differed in gender ratio, (4) knowledge of the grading system before entry into medical school may have altered the composition of the graded and pass/fail classes in some manner, (5) we do not know whether the two medical classes used in this study entered medical school with the same or a different average level of psychological

well-being, (6) medical students were aware that a research study on the effects of grading was under way, and (7) the results of the study can best be generalized to medical schools with a similar student body and similar curriculum.

This study provides useful information for medical schools contemplating a change in grading system in the early years of medical school from a graded (A, B, C, D, F) to a pass/fail grading system as to the probable impact on medical student performance, attendance, residency placement, satisfaction with medical school, and psychological well-being. Similarly, it provides reassurance for medical schools that have already chosen to make such a change.

Acknowledgments

References

1 Moss TJ, Deland EC, Maloney JV. Selection of medical students for graduate training: Pass/fail versus grades. N Engl J Med. 1978; 299:25–27.
2 Robins LS, Fantone JC, Oh MS, Alexander GL, Shlafer M, Davis WK. The effect of pass/fail grading and weekly quizzes on first-year students' performances and satisfaction. Acad Med. 1995;70:327–329.
3 Vosti KL, Jacobs CD. Outcome measurements in postgraduate year one of graduates from a medical school with a pass/fail grading system. Acad Med. 1999;74:547–549.
4 Jones KG, Pedersen RL, Carmichael SW, Pawlina W. Effects of pass/fail grading system on academic performance of first year medical students in gross anatomy course [abstract]. FASEB J. 2003;17:A385. Abstract 278.18.
5 Rohe DE, Barrier PA, Clark ME, Cook DA, Vickers KS, Decker PA. The benefits of pass–fail grading on stress, mood and group cohesion in medical students. Mayo Clin Proc. 2006;81:1443–1448.
6 Wolf TM. Stress, coping and health: Enhancing well-being during medical school. Med Educ. 1994;28:8–17.
7 Radcliffe C, Lester H. Perceived stress during undergraduate medical training: A qualitative study. Med Educ. 2003;37:32–38.
8 Rosenthal TL, Edwards NB, Ackerman BJ. Students' self-ratings of subjective stress across 30 months of medical school. Behav Res Ther. 1987;25:155–158.
9 Dupuy HI, McDowell I, Newell C. The general well-being schedule. In: Measuring Health: A Guide to Rating Scales and Questionnaires. 2nd ed. New York, NY: Oxford University Press; 1996;206–213.
10 Fazio AF. A Concurrent Validation Study of the NCHS General Well-Being Schedule. Hyattsville, Md: U.S. Department of Health, Education and Welfare, National Center for Health Statistics; 1977. Vital and Health Statistics Series 2, No. 73. DHEW Publication No. (HRA) 78-1347.
11 Dupuy HJ. The Psychological General Well-Being (PGWB) index. In: Wenger NK, Mattson ME, Furberg CD, Elinson J, eds. Assessment of Quality of Life in Clinical Trials of Cardiovascular Disease. New York, NY: Le Jacq Publishing, Inc.; 1984;170–183.
12 Hughes RL, Golmon ME, Patterson R. The grading system as a factor in the selection of residents. J Med Educ. 1983;58:479–481.
13 Kidson M, Hornblow A. Examination anxiety in medical students: Experiences with the visual analogue scale for anxiety. Med Educ. 1982;16:247–250.
14 Wolf TM, Faucett JM, Randall HM, Balson PM. Graduating medical students' ratings of stresses, pleasures, and coping strategies. J Med Educ. 1988;63:636–642.
15 Clark DC, Zeldow PB. Vicissitudes of depressed mood during four years of medical school. JAMA. 1988;260:2521–2528.
16 Wagoner NE, Suriano JR, Stoner J. Factors used by program directors to select residents. J Med Educ. 1986;61:10–21.
17 Wagoner NE, Suriano JR. Program directors' responses to a survey on variables used to select residents in a time of change. Acad Med. 1999;74:51–58.
18 Poirier MP, Pruitt CW. Factors used by pediatric emergency medicine program directors to select their fellows. Pediatr Emerg Care. 2003;19:157–161.
19 Crane JT, Ferraro CM. Selection criteria for emergency medicine residency applicants. Acad Emerg Med. 2000;7:54–60.
20 Tardiff K. The effect of pass–fail on the selection and performance of residents. J Med Educ. 1980;55:656–661.