The Benefits of Pass-Fail Grading on Stress, Mood, and Group Cohesion in Medical Students

Daniel E. Rohe, PhD; Patricia A. Barrier, MD; Matthew M. Clark, PhD; David A. Cook, MD; Kristin S. Vickers, PhD; and Paul A. Decker, MS

OBJECTIVE: To objectively measure the effect of a pass-fail grading system on stress, mood, group cohesion, and test anxiety in medical students.

PARTICIPANTS AND METHODS: Beginning with the class of 2006, the Mayo Medical School, Rochester, Minn, changed the grading system for first-year courses from a 5-interval grading system to a pass-fail grading system. Students in the previous class of 2005, who were graded using a 5-interval system during their first year of medical school, were compared with students in the class of 2006. Using a prospective study design, the 2 groups were compared at the end of both the first year and the second year of medical school on the Perceived Stress Scale, Profile of Mood States, Perceived Cohesion Scale, Test Anxiety Inventory, and (after year 2) the United States Medical Licensing Examination Step 1. Data collection occurred in 2002 and 2003 with the class of 2005 and in 2003 and 2004 with the class of 2006.

RESULTS: Students graded with the pass-fail system had less perceived stress (median, 15.0 vs 21.0; P = .01) and greater group cohesion (median, 34.5 vs 30.0; P = .02) at the end of their second year of coursework than their 5-interval graded peers. The pass-fail group had better mood (median, 46.5) than the graded group (median, 64.0), but this difference was not statistically significant (P = .07). No significant differences were found between the 2 groups in test-taking anxiety or in United States Medical Licensing Examination Step 1 board scores.

CONCLUSION: Pass-fail grading may reduce stress and increase group cohesion in medical students compared with traditional 5-interval grading.


Medical student distress is an area of increasing concern. A recent review of the literature by Dyrbye et al describes the manifestations, causes, and consequences of medical student distress. They note that some aspects of medical education may have unintended negative effects on medical students’ mental and emotional health. They specifically cite the 5-level, A through F, grading system as creating a competitive environment that promotes anxiety and peer competition rather than collaborative learning. Forsyth2 noted that letter grades represent massive extrinsic rewards, which by their nature tend to transform intrinsically motivated learners into extrinsically motivated learners, precisely the opposite of currently espoused medical education principles. These concerns raise the issue of whether the pass-fail grading system has quantifiable benefits compared with a 5-level grading system.

The early literature on pass-fail grading focused on the issues of academic achievement and student approval. In one of the earliest arguments in support of adopting a pass-fail grading system in medical schools, Miller1 stated that the traditional medical school grading system failed the 4 basic functions of grading and therefore was not consistent with professionalism or enhancing adult learning. A survey by Bender4 of medical school deans and student representatives found that a pass-fail system was desired by two thirds of the respondents. Robins et al5 compared a 4-interval system to a pass-fail system and found no difference in academic achievement. In a study of residents in their first postgraduate year, Vosti and Jacobs6 showed that graduates of schools with a pass-fail grading system performed similarly to peers who graduated from schools that used a 4-interval grading system. Jones et al7 found that changing from a letter to a pass-fail grading system in the first-year anatomy course was favored by students and had no effect on the National Board of Medical Examiners subject examination or final cumulative grades. However, not everyone is convinced that the benefits of a pass-fail grading system outweigh the benefits of the traditional letter grading system.8,9 For example, Moss et al10 analyzed surgical residents in their program who came from schools that had either pass-fail or letter grading systems and found that surgical residents from a pass-fail grading system accounted for 82% of those ranking below the 15th percentile and that none ranked above the 87th percentile of their surgical residents. Gonnella et al11 found substantial correlation between number grades from year 1 and subsequent medical school academic performance, medical licensing
examination scores, and clinical competence ratings in the first postgraduate year. These 2 studies suggest that numerical distinctions may provide important information related to future academic performance. However, 2 longitudinal studies found no significant relationship between grade point average in medical school and subsequent performance in medical practice.\textsuperscript{12,13}

Although research analyzing academic achievement and grading systems suggests little, if any, harm from the pass-fail system, less is known about the benefits of pass-fail grading to students. In theory, these benefits would derive from improvements in the psychological well-being of students, thus enhancing the learning environment by promoting cooperation and group learning.\textsuperscript{14,15} However, little research has been published on the psychological impact of changing to a pass-fail system in medical schools. Robins et al\textsuperscript{11} used student questionnaires and found that a pass-fail system appeared to reduce anxiety and within-class competition. In an unpublished survey of US medical school deans conducted by Mayo Medical School, Rochester, Minn, several anecdotal benefits in changing to a pass-fail grading system were noted. These deans perceived decreased student anxiety, class competitiveness, and grade anxiety and increased student collaboration (A. J. Windebank, MD, written communication, July 31, 2002). This research is limited because outcomes were derived from informal student self-reports and dean observation rather than objective measurement. Furthermore, no published studies have evaluated the long-term psychological effect of a pass-fail system. Although the immediate effect on group learning may be important, the most powerful effect may be more durable, with positive affective changes achieved early in the curriculum that carry over into subsequent years.

Beginning with the class of 2006, the Mayo Medical School replaced the 5-interval grading system with a pass-marginal pass-fail system for the first year of medical school. Although not a strictly pass-fail system, the new grading system requires students to remediate any marginal pass grade; thus, it is in effect a pass-fail system. This change in grading system provided the opportunity to prospectively investigate changes in student well-being between the 2 grading systems.

The purpose of the current study was to determine the potential sustained (primary aim) and immediate (secondary aim) effects of a pass-fail grading system in the first year of medical school on stress, mood, group cohesion, and test anxiety compared with a 5-level grading system. To accomplish this, we conducted a prospective static-group, posttest-only study comparing 2 consecutive first-year classes and assessing these outcomes using standardized measures at the end of the first academic year and again at the end of the second year. We hypothesized that medical students graded with use of a pass-fail system would exhibit less stress, improved mood, greater group cohesion, and less test anxiety compared with their colleagues graded using a 5-interval scale at the end of their first year of medical school. We further hypothesized that these effects would persist at the end of the second year despite the grading system having reverted to a 5-level system for those initially graded using a 2-level system during their first year of medical school. Data collected at the end of the second year were selected as the primary end point for 2 reasons. First, the data were gathered several weeks after subject examinations were finished, which minimized the likelihood that the response to the instruments would reflect a transient increase in distress during subject examinations. Second, we were primarily interested in whether the potential benefits of the pass-fail grading system during the first year would continue into the second year of medical school even though the grading system had reverted to the 5-interval system.

PARTICIPANTS AND METHODS

All Mayo Medical School students in the graduating classes of 2005 and 2006 were eligible to participate. The study was reviewed and approved by the Mayo Foundation Institutional Review Board. Participation was voluntary, and written informed consent was obtained from all participants.

DESIGN AND PROCEDURE

The Mayo Medical School class of 2005, designated the graded group, and the Mayo Medical School class of 2006, designated the pass-fail group, completed all 4 study instruments at the end of their first year of medical school (time 1) and again at the end of their second year of medical school (time 2). Data collection occurred in 2002 and 2003 with the class of 2005 and in 2003 and 2004 with the class of 2006. At time 2, the graded group had completed 2 years of medical school with the 5-interval grading system in effect. At time 2, the pass-fail group had completed their first year of medical school under the pass-fail system and their second year of medical school under the 5-interval system.

OUTCOME MEASURES

The 4 self-report instruments used in the study were the Perceived Stress Scale (PSS),\textsuperscript{16} Profile of Mood States (POMS),\textsuperscript{17} Perceived Cohesion Scale (PCS),\textsuperscript{18} and Test Attitude Inventory (TAI).\textsuperscript{19} A secondary outcome, academic performance, was assessed using the United States Medical Licensing Examination (USMLE; www.nbme.org) Step 1.

The PSS is a 10-item, Likert-scale self-report instrument designed to measure the extent to which situations in
an individual’s life are appraised as stressful during the past month (lower scores indicate less stress). The PSS scores have an internal consistency of 0.85, a test-retest reliability of 0.85, and a concurrent and predictive validity in samples of college students and adults.16

The POMS is a 65-item, factor analytically derived adjective rating scale that measures 6 affective states (tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment) that when added produce a total mood disturbance score (lower scores indicate less disturbed mood). Respondents use a 5-point Likert scale to rate the degree to which each adjective describes how they have been feeling in the past week. Test-retest reliabilities for scores representing the 6 factors ranged from 0.61 to 0.69 throughout 4 weeks.17

The PCS is a 6-item, Likert-scale self-report instrument designed to measure an individual’s perception of cohesion within a group, defined as “an individual’s sense of belonging to a particular group and his or her feelings of morale associated with membership in the group.”18 Higher scores indicate more cohesion. The revised form of the PCS, following the recommendations of Bollen and Hoyle,18 was used. Bollen and Hoyle18 found that the instrument had 2 underlying factors, entitled belongingness and morale, with a Cronbach α of .97 and .87, respectively.20

The TAI is a 20-item, Likert-scale self-report instrument that measures individual differences in test anxiety as a situation-specific personality trait. The instrument measures anxiety proneness in test situations, with lower scores indicating less anxiety. Test-retest reliability coefficients range from 0.80 for 1 week to 0.62 for 6 weeks.19

### Statistical Analyses

Baseline demographics were compared between medical school classes using the 2-sample Wilcoxon rank sum test for continuous variables and the χ² test for categorical variables. Scores on the PSS, POMS, PCS, and TAI were compared using the 2-sample Wilcoxon rank sum test. This analysis was conducted for both time 1 and time 2. Additionally, the USMLE Step 1 test scores were compared using the 2-sample t test. In all cases, 2-sided P≤.05 was considered statistically significant.

### RESULTS

In the graded group, all 41 students consented to participate at time 1, and 37 (90%) completed instruments at time 2. In the pass-fail group, 39 (98%) of 40 consented to participate at time 1, and 38 (95%) completed instruments at time 2. The number of participants who completed each inventory is listed in the respective tables. To protect participant identity, we did not collect demographic information on individual participants and instead report demographic information for both entire medical school classes. Given that all of the 2005 class and all but 1 student in the 2006 class participated at time 1, this information appears to be representative of the study participants. No statistically significant differences were found between the 2 classes on sex, Medical College Admission Test scores, or grade point averages (Table 1). The class of 2006 was older than the class of 2005 (mean age, 24.3 vs 22.7 years; P=.01).

Table 2 lists scores from the PSS, POMS, PCS, and TAI after the first year of medical school. These data reveal that at the end of their first year of medical school, those students who were evaluated with a pass-fail grading system rated themselves as having significantly less stress, better overall mood, and greater group cohesion compared with their letter-graded peers. The pass-fail group’s score on the TAI was lower than their letter-graded peers, but this difference was not statistically significant (P=.18).

| TABLE 1. Demographics of First-Year Students in Classes of 2005 and 2006* |
|-----------------------------|-----------------------------|-----------------------------|
| Demographics                | Class of 2005 (n=41) | Class of 2006 (n=40) | P value† |
| No. (%) of participants at time 1 | (5-interval grading) | (pass-fail grading) | |
| No. (%) of participants at time 2 | 41 (100) | 39 (98) | |
| Mean (SD) age (y)‡ | 22.7 (1.7) | 24.3 (3.3) | .01 |
| No. (%) female‡ | 26 (63) | 20 (50) | .22 |
| Mean (SD) MCAT score‡ | 33.0 (3.7) | 32.6 (3.4) | .44 |
| Mean (SD) undergraduate GPA‡ | 3.8 (0.3) | 3.8 (0.2) | .86 |

*GPA = grade point average; MCAT = Medical College Admission Test.
†Two-sample Wilcoxon rank sum test or χ² test as appropriate.
‡These demographics represent the entire classes and not the specific students who consented to participate.
BENEFITS OF PASS-FAIL GRADING

The main focus of this investigation was to examine potential differences between the 2 classes at the end of their second year of medical school. Table 3 contains scores from the 4 study instruments. It is worth reiterating that the pass-fail group shifted from pass-fail grading in their first year to a 5-level grading system in their second year. The data in Table 3 indicate that at the end of their second year of medical school, students who were graded using a pass-fail grading system during their first year of medical school continued to perceive less stress and greater group cohesion than their letter-graded peers. There was a trend for the pass-fail group to have better mood (median, 46.5 vs 64.0; \( P = .07 \)). Again, no difference was found between the groups in test anxiety. The USMLE Step 1 scores were not statistically significantly different between the 2 groups (mean \( \pm \) SD for graded group vs pass-fail group, 234 \( \pm \) 22 vs 227 \( \pm \) 21; \( P = .15 \)).

### DISCUSSION

To our knowledge, this is the first prospective study to report on the psychological effect of using a pass-fail grading system during the first year of medical school. Results indicate that the students experienced less perceived stress, better mood, and greater group cohesiveness among a cohort of medical students using a pass-fail grading system during their first year of medical school compared with a separate cohort graded using the traditional 5-interval system. More importantly, these benefits persisted to the end of the second year of medical school, even though grading reverted to the traditional 5-level system for the second year. No apparent difference was found in academic performance between the 2 groups. These results provide the first objective evidence supporting the psychological benefits of the pass-fail system over the traditional 5-interval grading system during the basic science years of an undergraduate medical school curriculum.

These data corroborate the subjective impression of medical school deans whose medical schools switched to pass-fail grading systems. These deans observed that with a pass-fail grading system their students are less stressed, have a greater positive mood, have greater group cohesion (less within-group competition), and performed no differently on objective measures of scholastic performance compared with students for whom the 5-interval grading system was used. Data from the current study are also congruent with the findings of Robins et al \(^5\) that students reported reduced anxiety and within-class competition and

### TABLE 2. Graded Group vs Pass-Fail Group Test Scores at the End of the First Year of Medical School*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Graded group</th>
<th>Pass-fail group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.† Median</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td>41 13.0 13.8 (6.4)</td>
<td>38 9.0 10.9 (6.2)</td>
</tr>
<tr>
<td>Profile of Mood States</td>
<td>39 23.0 32.0 (39.0)</td>
<td>38 11.0 13.0 (23.5)</td>
</tr>
<tr>
<td>Perceived Cohesion Scale</td>
<td>41 35.0 32.9 (8.4)</td>
<td>39 39.0 37.8 (5.5)</td>
</tr>
<tr>
<td>Test Attitude Inventory</td>
<td>40 33.0 36.7 (14.3)</td>
<td>39 30.0 31.7 (9.3)</td>
</tr>
</tbody>
</table>

*For the Perceived Stress Scale, Profile of Mood States, Perceived Cohesion Scale, and Test Attitude Inventory, lower scores indicate less of that psychological state (ie, less stress, improved mood, less cohesion, and lower test anxiety).

†All students in the graded group and 39 of 40 in the pass-fail group consented to participate. In several instances, not all instruments were completed correctly and data were not usable.

‡Two-sample Wilcoxon rank sum test comparing the graded group to the pass-fail group.

### TABLE 3. Graded Group vs Pass-Fail Group Test Scores at the End of the Second Year of Medical School*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Graded group</th>
<th>Pass-fail group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.† Median</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td>37 21.0 20.5 (7.8)</td>
<td>38 15.0 15.8 (6.8)</td>
</tr>
<tr>
<td>Profile of Mood States</td>
<td>34 64.0 64.6 (40.5)</td>
<td>38 46.5 47.1 (31.9)</td>
</tr>
<tr>
<td>Perceived Cohesion Scale</td>
<td>37 30.0 29.0 (9.9)</td>
<td>38 34.5 33.8 (8.0)</td>
</tr>
<tr>
<td>Test Attitude Inventory</td>
<td>37 36.0 38.3 (13.8)</td>
<td>37 37.0 37.2 (12.6)</td>
</tr>
</tbody>
</table>

*For the Perceived Stress Scale, Profile of Mood States, Perceived Cohesion Scale, and Test Attitude Inventory, lower scores indicate less of that psychological state (ie, less stress, improved mood, less cohesion, and lower test anxiety).

†All students in the graded group and 39 of 40 in the pass-fail group consented to participate. In several instances, not all instruments were completed correctly and data were not usable.

‡Two-sample Wilcoxon rank sum test comparing the graded group with the pass-fail group.
greater student cooperation with a pass-fail grading system. Although the current study did not directly assess competition, we believe reduced group competition can be logically inferred on the basis of higher levels of group cohesion. These results have implications for medical education reform, medical student well-being, and student evaluation.

The evaluation system used is a direct reflection of the educator’s paradigm of teaching. The paradigm of teaching in turn determines how education is structured. Research by Johnson and colleagues on the old vs the new paradigm of education is particularly relevant to medical education reform. The old paradigm places students in a dependent, passive, and competitive role and the faculty in a sorting and classifying role. The new paradigm places students in an independent, active, and cooperative role and the faculty in a developing competencies and talents role. Johnson and colleagues state that educators have a choice of structuring every learning task competitively, cooperatively, or individually.

In competitive learning situations, individuals seek outcomes that are beneficial to themselves and detrimental to others. In cooperative learning situations, individuals seek outcomes that are beneficial to themselves and all other group members. Students enter medical school by virtue of success in an intensely competitive learning environment. Success as a physician, however, requires the replacement of a competitive mindset with its opposite, a cooperative mindset. A cooperative mindset facilitates building relationships within the health care team. Working effectively in multidisciplinary teams is a competency specifically mandated of undergraduate and graduate medical school students. The skills required to successfully work in multidisciplinary health care teams presuppose a positive attitude toward cooperation. We advocate that use of the pass-fail grading system is congruent with the new paradigm of education and encourages educators to structure learning tasks cooperatively. In a related topic, Stagnaro-Green documents the ongoing failure of medical education to incorporate the principles of adult education. A potential explanation for this failure may be the impact of a 5-interval grading system on student and teacher behavior.

An established principle in education is that the first learning is the most powerful learning. Our data suggest that the grading system used has the power to reduce stress, foster positive mood, and enhance a sense of group cohesion in medical students. By using a stress-minimizing evaluation system at the beginning of the medical student’s educational experience, a message of respect and concern will be communicated. An evaluation system that reflects respect and concern for the student may potentially affect attitudes of concern for future patients.

Regarding the interplay between student stress and learning, Johnson et al state, “There is a general rule of instruction: The more pressure placed on students to achieve and the more difficult the material to be learned, the more important it is to provide social support within the learning situation. Challenge and support must be balanced if students are to cope successfully with the stress inherent in learning situations.”

As suggested by Dyrbye et al, nowhere is the interplay among stress, difficulty of the material, and need for social support more apparent than in medical education. The available research shows that cooperative learning results in greater effort to achieve, more positive interpersonal relationships, and greater psychological health and self-esteem than does competitive and individualistic learning. Our data suggest that the pass-fail grading system may lead to enhanced group cohesion and cooperation.

The current study has several limitations. First, since this was a nonrandomized study, the 2 cohorts were different in age. Although both cohorts were composed of young adults, perhaps age, as an indicator of maturity, influenced the findings. Additionally, other differences may have occurred between the 2 classes at entry into medical school that may explain the observed differences. Unfortunately, baseline data on stress, cohesiveness, mood, or test anxiety that might have helped address this weakness were not collected. Also, demographic information was available only for the entire class and not for individual participants; however, given the high rate of study participation, the class-wide demographics are a reasonable approximation of the study participants. Second, we cannot exclude the possibility that these differences are the product of factors within or external to the school curriculum yet unrelated to the method of grading. Third, the aims of the study were transparent and may have led students to consciously or unconsciously respond to the instruments in a manner that would help to support the pass-fail grading system. Fourth, the students at our medical school may not be representative of those from other schools. Fifth, our pass-fail grading system had an intermediate level, marginal pass, which limits the direct extrapolation to purely 2-level grading systems. Finally, the nonsignificant difference in the USMLE Step 1 scores may have been due to small sample size.

Future research should be conducted to replicate and extend our findings. Studies should also be performed to establish a causal link between the grading system and the outcomes we assessed, investigate the impact of the pass-fail system on other outcomes such as multidisciplinary teamwork, and evaluate the long-term impact of the pass-fail system on such outcomes as reducing cynicism and improving quality of patient care.
CONCLUSION

This study provides preliminary evidence supporting the benefit of a pass-fail grading system on psychological well-being, including decreased stress, improved mood, and increased group cohesiveness in medical school students. These data also suggest that the positive influence of a pass-fail system during the first year of medical school persists even after reverting to a traditional 5-level grading system in the second year of medical school. This study goes beyond previous research, which demonstrated the absence of harm, and provides positive reasons that support increased use of the pass-fail grading system.

REFERENCES