Pass–fail grading: laying the foundation for self-regulated learning

Casey B. White · Joseph C. Fantone

Abstract Traditionally, medical schools have tended to make assumptions that students will "automatically" engage in self-education effectively after graduation and subsequent training in residency and fellowships. In reality, the majority of medical graduates out in practice feel unprepared for learning on their own. Many medical schools are now adopting strategies and pedagogies to help students become self-regulating learners. Along with these changes in practices and pedagogy, many schools are eliminating a cornerstone of extrinsic motivation: discriminating grades. To study the effects of the switch from discriminating to pass–fail grading in the second year of medical school, we compared internal and external assessments and evaluations for a second-year class with a discriminating grading scale (Honors, High Pass, Pass, Fail) and for a second-year class with a pass–fail grading scale. Of the measures we compared (MCATs, GPAs, means on second-year examinations, USMLE Step 1 scores, residency placement, in which there were no statistically significant changes), the only statistically significant decreases (lower performance with pass fail) were found in two of the second-year courses. Performance in one other course also improved significantly. Pass–fail grading can meet several important intended outcomes, including "leveling the playing field" for incoming students with different academic backgrounds, reducing competition and fostering collaboration among members of a class, more time for extracurricular interests and personal activities. Pass–fail grading also reduces competition and supports collaboration, and fosters intrinsic motivation, which is key to self-regulated, lifelong learning.

Keywords Assessment · Collaborative learning · Grading · Pass/fail grading · Self-regulated learning

Physicians in practice are expected to maintain their competence throughout their careers by participating in continuous learning processes. To guide their learning, they self-monitor and set learning goals, and then address their gaps or acquire new knowledge and

C. B. White (✉) · J. C. Fantone
University of Michigan Medical School, 1135 Catherine Street, Ann Arbor, MI 48109, USA
e-mail: bcwhite@umich.edu

skills through a range of educational activities. This sequence of activities comprises self-regulated learning (Zimmerman 2000), which, although it underlies safe and effective medical practice, is not intentionally taught or monitored in most medical education settings. Traditionally, medical schools have tended to make assumptions that students will "automatically" engage in self-education effectively after graduation and subsequent training in residency and fellowships. In reality, the majority of medical graduates out in practice feel unprepared for learning on their own (Weed 1997).

Importantly, we now have evidence that self-regulated learning (SRL) can be modeled and taught in educational settings (Schunk and Zimmerman 1998; Pintrich 1994). A key element in SRL is motivation, which has been typically discussed as either intrinsic or extrinsic (White 2007). In this context, intrinsic motivation is described as students engaging in "deep" learning, based on interest and enjoyment. Extrinsic motivation is described as students engaging in "surface" learning, based on external rewards such as test scores or course grades (Deci 1992; Vansteenkiste et al. 2004). Typically, effective self-regulated learners tend to be more intrinsically than extrinsically motivated (Pintrich and DeGroot 1990).

Many medical schools are now adopting strategies and pedagogies to help students become self-regulating learners, including more active, collaborative and student-centered methods, with expectations (sometimes explicit and sometimes not) that students have a clear role in their own learning and achievement of institutional and personal goals. Along with these changes in practices and pedagogy, more schools than before[1] have eliminated a cornerstone of extrinsic motivation: discriminating grades.

Those who are more supportive of a discriminating grading system see it as a motivator; they are concerned that without this motivator academic performance will decline (or poor performance will be "covered up") and residency placement will be compromised (Dietrick et al. 1991; Gonella 2004; Gutman and Grufferman 1985; Provan 1995). Those more supportive of a pass–fail grading system believe that performance discrimination inhibits deeper learning (i.e., critical analysis of new ideas; connecting those ideas to already known concepts and principles) that is linked with intrinsic motivation, and encourages competition that hinders effective collaborative learning (Phillips 1965; Robins et al. 1995; Weller 1983). Discriminating grading thus directly conflicts with approaches that support self-regulated learning. Pass–fail supporters further contend that pass–fail grading provides an environment that encourages deeper learning in areas of interest to students, and does *not* diminish overall academic performance (Rohe et al. 2006; Vosti and Jacobs 1999) or compromise residency placement (Vosti and Jacobs 1999). In fact, after adopting a pass–fail grading scheme, at least one other study also found an increase in student satisfaction and improved well-being (Bloodgood et al. 2009). In 1992 we adopted a pass–fail grading scale in the first year of the medical student curriculum. We studied its short-term effects and found that performance—usually quite high—was unchanged, and students were more satisfied and less competitive (Robins et al. 1995). In 2003 we implemented a new curriculum that integrated pedagogical elements specifically designed to foster SRL (developing personal learning goals, reflecting on effective learning strategies, accountability to self and peers for learning, self-monitoring against goals). Grading in the first year was still Pass–Fail (Pass = ≥75%) and grading in the second year was still discriminating (92 = 100% = Honors, 86–91.9% = High Pass, 70–85.9% = Pass, <70% = Fail). Two years later we determined that the discriminating grading scale in the second year was undermining our efforts to promote peer teaching and collaborative learning

---

[1] AAMC website http://services.aamc.org/currdir/section1/grading1.cfm

(i.e., students who compete for grades do not collaborate effectively on learning) (Kohn 1986; Smith et al. 2005). To support pedagogical efforts designed to lay a foundation for effective SRL, the faculty ultimately reached consensus and approved a pass–fail grading scale in the second year beginning in 2005–2006.

In this study we explored the effects of pass–fail grading in the second year of medical school. We compared a class in our new curriculum that had a discriminating grading scale in the second year (2004–2005) with a class in the same curriculum that had a pass–fail grading scale in the second year (2005–2006). Specific areas we investigated to see if performance or outcomes changed included:

1. Performance on second year examinations.
2. Performance on USMLE Step 1.
3. Performance on USMLE Step 2.
4. Student satisfaction with pass–fail grading.
5. Residency placement.

Methods

To study the effects of pass–fail grading in the second year of medical school, we reviewed and compared a number of internal and external assessments and evaluations for a second-year class with a discriminating grading scale (Honors, High Pass, Pass, Fail) and for a second-year class with a pass–fail grading scale.

Admissions credentials of the two classes, including MCAT average scores and average grade point averages for the two classes, were compared using independent samples t-tests to determine if there were statistically significant differences in performance between the two cohorts of students. Both classes had identical medical school courses, experiences and requirements.

Cumulative exam scores from the eleven second-year basic science courses were compared using independent-samples t-tests to determine if there were statistically significant differences in performance between the two cohorts of students. Similarly, US-MLE Step 1 and Step 2 scores for the two groups were compared using t-tests.

To examine student satisfaction with the discriminating scale compared with pass–fail grading, we used response data from annual curriculum evaluation surveys. At the end of each year of medical school, students complete a comprehensive survey of their curriculum and learning environment experiences. On the second-year survey in 2004–2005, students rated their "Satisfaction with traditional grading in year 2" on a 5-point Likert scale where 1 = Not Satisfied At All and 5 = Exceptionally Satisfied. On the second-year survey in 2005–2006, students rated their "Satisfaction with pass–fail grading in year 2" on the same 5-point scale. Independent-samples t-tests were used to determine if there were statistically significant differences in student satisfaction between the two groups.

On the same survey, in 2005–2006 students were asked to rate five separate items related to pass–fail grading on a 4-point scale where 1 = Strongly Disagree and 4 = Strongly Agree. These items were designed to explore whether and to what extent pass/fail grading allowed students to engage in other areas of interest.

Two related measures were used to compare the two groups of students in terms of residency placement. National Residency Match Program (NRMP) results for the class that graduated in 2007 (second year 2004/2005) and the class that graduated in 2008 (second year 2005/2006) were sent to each of the residency program directors at our medical

school. We asked them to rank each residency program in their specialty area on the match list on a scale where 1 = Top 15 US program, 2 = Top 30 US program, and 3 = Other. We utilized paired-samples t-tests to compare mean rankings by residency directors to the two cohorts' residency matches, and Fisher's exact test to compare the proportion of students in each group who were placed in a Top 15 ranked program.

Results

There were no statistically significant differences in admissions credentials for the second year class in 2004/2005 versus the second year class in 2005/2006. MCAT average was 11.48 for the 2004/2005 versus 11.40 for the 2005/2006 class ($p = 0.806$); GPA average was 3.75 for the 2004/2005 class versus 3.74 for the 2005/2006 class ($p = 0.569$).

There were statistically significant differences in performance in three of the second year sequences between 2004/2005 and 2005/2006 (we defined statistical significance as $p < 0.0045$, following a Bonferroni adjustment of 0.05/11, for the number of comparisons made). Performance declined in two of the sequences: Cardiology (94.0% in 2004/2005, 90.3% in 2005/2006, $p = 0.000$), and Reproduction (93.9% in 2004/2005, 89.1% in 2005/2006, $p = 0.000$). Performance improved in the Renal sequence (89.4% in 2004/2005, 92.2% in 2005/2006, $p = 0.000$; Table 1).

There were no statistically significant differences in performance on USMLE Step 1 (231 for second-year class in 2004/2005, 232 for second-year class in 2005/2006, $p = 0.628$) or on USMLE Step 2 (239 for second-year class in 2004/2005, 241 for second-year class in 2005/2006, $p = 0.377$; Table 2).

Both second-year classes were similarly satisfied with pass–fail grading in the first-year curriculum. Second-year students in 2004/2005 rated pass–fail 4.76 on a 5-point Likert scale; second-year students in 2005/2006 rated it 4.77. Second-year students in 2004/2005 rated traditional grading in year two 2.96 on the same scale, while second-year students in 2005/2006 rated pass–fail grading in year two 4.64, a significantly higher mean than the students who had traditional grading in year two (Table 3).

Members of the 2005/2006 second-year class were also asked to respond to additional items related to pass–fail grading in year two. On a 4-point scale where 1 = strongly disagree and 4 = strongly agree, this class rated a series of items designed to explore whether pass–fail grading freed up study time, and how they used that time; responses ranged from 3.07 (more time to explore additional academic talents) to 3.43 (more time to exercise and improve personal wellness; Table 4).

We asked residency directors to rank each of the programs into which the two classes (second year in 2004/2005 and second year in 2005/2006) had matched, using a 3-point scale where 1 = Top 15 program in that discipline, 2 = Top 30 in that discipline, and 3 = Other. For the class that had completed the second year in 2004/2005 (traditional grading), the average ranking across students and disciplines was 1.68 (SD = 0.80); for the class that had completed the second year in 2005/2006 the average ranking was 1.70 (0.78; Table 5). There was no statistically significant difference between the two classes.

We also calculated how many students in each of the two classes had matched into a Top 15 program. For the class that had completed the second year in 2004/2005 that number was 54 (30% of the class); for the class that had completed the second year in 2005/2006 that number was 50 (32% of the class). There was no statistically significant difference between the two classes.

Table 1 Average cumulative scores on second-year coursework 2004–2005 and 2005–2006 (listed in order taken)

	2004–2005 mean (SD), N	2005–2006 mean (SD), N	Cohen's d (effect size)**	Student's t	SE diff.	2-tail p-value
Cardiology*	94.0% (3.91) n = 169	90.3% (4.14) n = 171	0.92	8.49	0.437	0.000
Respiratory	94.1% (4.00) n = 169	93.4% (4.04) n = 171	0.17	1.58	0.436	0.115
Renal*	89.4% (6.53) n = 169	92.2% (6.01) n = 171	0.44	4.04	0.681	0.000
Psychiatry	93.5% (4.51) n = 169	92.9% (5.98) n = 172	0.12	1.10	0.574	0.274
Neurology	86.4% (6.54) n = 169	87.4% (5.33) n = 173	0.17	1.57	0.644	0.118
Musculoskeletal	91.4% (6.99) n = 169	90.2% (5.09) n = 173	0.19	1.79	0.660	0.075
Dermatology	95.2% (3.59) n = 168	94.9% (2.63) n = 170	0.10	0.91	0.342	0.365
Hematology	91.0% (6.17) n = 166	90.9% (6.64) n = 172	0.02	0.16	0.698	0.875
Gastroenterology	88.3% (5.12) n = 168	88.0% (4.58) n = 171	0.06	0.57	0.527	0.570
Endocrine	91.0% (5.18) n = 167	89.8% (4.95) n = 173	0.24	2.24	0.549	0.026
Reproduction*	93.9% (4.97) n = 167	89.1% (4.25) n = 173	1.05	9.70	0.501	0.000

* Following a Bonferroni adjustment for the number of comparisons, these results, where $p < 0.0045$, were considered statistically significant

** By convention, values of Cohen's d may be interpreted as follows: 0.20 or less indicate a small difference, values in the range of 0.50 indicate a moderate difference, and values of 0.80 or greater indicate a large difference

Table 2 USMLE step 1 and step 2 averages for second-year classes 2004/2005 and 2005/2006

	2004–2005 mean (SD), N	2005–2006 mean (SD), N	Cohen's d (effect size)*	Student's t	SE diff.	2-tail p-value
USMLE Step 1	231 (19) n = 165	232 (19) n = 174	0.05	0.48	2.065	0.628
USMLE Step 2	239 (21) n = 160	241 (20) n = 169	0.10	0.88	2.260	0.377

* By convention, values of Cohen's d may be interpreted as follows: 0.20 or less indicate a small difference, values in the range of 0.50 indicate a moderate difference, and values of 0.80 or greater indicate a large difference

Discussion

Across levels and disciplines, few teachers or programs have integrated concrete elements designed to prepare students effectively for lifelong self-education (Zimmerman et al. 1996). Although medical schools "talk the talk", beyond problem-based learning, which focuses learning through clinical problems and on self-accountability and self-responsibility for learning, concrete changes in this direction have been few (White 2007), and reported outcomes limited.

Table 3 Student satisfaction with grading in year 1 and year 2 for second-year classes 2004/2005 and 2005/2006

	2004/2005	2005/2006
Satisfaction with pass–fail grading in year 1	4.75 (0.45) $n = 103$	4.77 (0.53) $n = 141$
Satisfaction with traditional grading in year 2	2.96 (1.20) $n = 103$	
Satisfaction with pass–fail grading in year 2		4.64 (0.71) $n = 141$

Based on a 5-point scale where 5 = exceptionally satisfied and 1 = not satisfied at all

Table 4 Student use of time freed up with pass–fail grading scheme for second-year class in 2005/2006

	2005/2006
Pass–Fail grading allowed me more time to explore additional academic talents	3.07 (0.75) $n = 124$
Pass–Fail grading allowed me more time to participate in volunteer/service activities	3.19 (0.66) $n = 135$
Pass–Fail grading allowed me more time to participate in student organizations	3.21 (0.70) $n = 137$
Pass–Fail grading allowed me more time to spend with my family	3.27 (0.69) $n = 131$
Pass–Fail grading allowed me more time to exercise and improve personal wellness	3.43 (0.59) $n = 137$

Based on a 4-point scale where 4 = strongly agree and 1 = strongly disagree

Table 5 Residency program directors' average ranking (SD) of program matches and placement in Top 15 of programs of students graduated in class of 2007 and class of 2008

	Class of 2007 (second year 2004/2005)	Class of 2008 (second year 2005/2006)	p value
Average rank by residency program directors*	1.68 (0.80) $n = 179$	1.70 (0.78) $n = 158$	Paired-sample t-test: 0.817
Number of class members placing in a Top 15 program	54 (30% of class)	50 (32% of class)	Fisher's exact: 0.814

* Based on a scale where 1 = Top 15 program, 2 = Top 30 program and 3 = Other

Extremely busy clinical faculty with restricted time to direct student learning have led to increased awareness among medical educators that medical students need better preparation to regulate their own learning as they begin clinical clerkships. The introduction of pass–fail grading in the second year of medical school was designed to allow students to establish individual goals for performance and monitor their performance against these goals, and to promote intrinsic motivation and de-emphasize extrinsic measures of success (grades). This constituted a formal step in linking—longitudinally—self-regulated learning earlier in medical school with self-regulated learning that is necessary in third- and fourth-year clerkships, and ultimately in residency and professional practice.

Some of the trappings of traditional medical student education—in particular discriminating grading scales that provide evidence of who is the "best of the best"—complicate matters. We know from previous studies that discriminating grading increases competition among students (White 2005), making it more difficult to introduce active, collaborative learning approaches successfully. Other concerns that have been expressed about replacing discriminating grades with pass–fail grading are effects on academic performance, satisfaction and residency placement; these needed to be addressed.

Although a specific goal of ours in adopting pass–fail grading was to promote intrinsic motivation, we are aware of evidence that links testing with increased extrinsic motivation (Phelps 2005). This might certainly be the case for some medical students. For instance, in the absence of discriminating grades that can be considered by potential residency programs in the match process, motivation to do well on the USMLE Step 1 might be increased. This might in turn motivate students to perform well on internal assessments, so they can measure their progress in learning what will be needed for the Step 1 examination. While this type of "performance motivation" does not necessarily conflict with our goal of fostering intrinsic motivation and self-regulation, there is research supporting the notion that there are different stages or levels of extrinsic motivation, some of which fit into its classic definition (doing something that is of value to someone else and leads to a separable outcome, such as a good grade or a gold star), and some of which indicates learners who are in active, autonomous states. Ryan and Deci described the most autonomous form of extrinsic motivation as *integration*, in which the value of an extrinsic goal is self-approved and adopted with some sense of having a choice (2000). Thus, in the context of pass–fail grading adopted to promote self-regulated learning, medical students are motivated to do well on internal tests because they value this measurement of their progress, and because they can choose how much time and effort to devote to learning. That is, they are traveling through the stages of extrinsic motivation toward the intrinsic motivation they will likely possess as professionals.

This study provides evidence that the introduction of pass–fail grading in the second year of medical school yields no differences in performance on external measures including residency placement. Internally, only one of 11 second-year courses reported a significant decrease in mean course performance greater than four points. This was the last course in the second year (Reproduction) and student feedback indicated that the decline was due to early preparation for the USMLE Step 1 examination, on which a passing score is a requirement for entry into the third year. With implementation of the proposed USMLE "Gateway" licensure examination in a few years (i.e., elimination of the Step 1 examination), this temporal distracter will likely no longer exist (Scoles 2008).

Students were very satisfied with the change to pass–fail grading in the second year. They reported additional time that they could devote to improving personal wellness, more time to explore extracurricular activities, and more time with family. Students had already been performing very highly on the examinations (Table 2). Given that the means did not drop across the courses with the introduction of pass–fail grading, and the students reported freed up time for other activities, we deduced that they had most likely been over-studying.

There have been concerns that pass–fail grading in the first 2 years of medical school lends inordinate significance to the USMLE Step 1 score as the only measure to discriminate among medical students' basic science knowledge in the residency application process. It is noteworthy that the compromise many medical schools have reached is to retain a discriminating grading scale in the clinical clerkships, and it is arguable that performance-based assessment in the clinical clerkships is a better predictor of performance in residency than performance in course work in years one and two. Further, as noted above, in the near future there is likely to be only one "Gateway" written licensure examination (along with a clinical skills assessment) administered during medical school, and the score on that assessment is likely to carry significant weight as the primary external measure of student learning across 3 years of medical school.

Conclusion

Pass–fail grading can meet several important intended outcomes, including "leveling the playing field" for incoming students with different academic backgrounds. An important planned outcome for us in adopting pass–fail grading is an effort to help our students become effective self-regulating learners. In this case, it is critical that they have a strong foundation in the cycle of self-regulated learning and its processes (White and Gruppen 2007). They must understand how their active participation in self-regulated learning activities during medical school, including setting personal goals and monitoring achievement, lays an important foundation for continuous learning throughout life.

Limitations

The literature on this subject area is varied, and much of it is not recent. Some of the previous studies explored pass–fail grading in the first year of medical school or the first 2 years of medical school, and others explored pass–fail grading versus discriminating grading in all 4 years of medical school.

University of Michigan medical students represent a selected sample of medical students in LCME-accredited medical schools.

This study has been approved by the University of Michigan Medical School Institutional Review Board.

References

Bloodgood, R. A., Short, J. G., Jackson, J. M., & Martindale, J. R. (2009). A change to pass-fail grading in the first two years at one medical school results in improved psychological well-being. *Academic Medicine, 84*(5), 655–662.

Deci, E. L. (1992). The relation of interest to the motivation of behavior: a self determination theory perspective. In K. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development*. Mawah, NJ: Lawrence Erlbaum.

Dietrick, J. A., Weaver, M. T., & Merrick, H. W. (1991). Pass-fail grading: A disadvantage for students applying for residency. *The American Journal of Surgery, 162*(1), 63–66.

Gonella, J. (2004). An empirical study of the predictive validity of number grades in medical school using 3 decades of longitudinal data: implications for a grading system. *Medical Education, 38*, 425–434.

Gutman, L., & Grufferman, S. (1985). Impact of medical school class ranking systems on applicants' ratings for residency positions. *Journal of Medical Education, 60*, 684–691.

Kohn, A. (1986). *No contest: The case against competition*. Boston, MA: Houghton Mifflin.

Phelps, R. P. (Ed.). (2005). *Defending standardized testing*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Phillips, J. (1965). An application of intermittent grading. *The Clearing House, 39*, 306.

Pintrich, P. (1994). *Handbook of self-regulated learning*. San Francisco, CA: Jossey-Bass.

Pintrich, P. R., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40.

Provan, J. (1995). Preferences of program directors for evaluation of candidates for postgraduate training. *Canadian Medical Association Journal, 153*(7), 919–923.

Robins, L., Fantone, J., Oh, M., Alexander, G., Shlafer, M., & Davis, W. K. (1995). The effect of pass-fail grading and weekly quizzes on first-year students' performances and satisfaction. *Academic Medicine, 70*(4), 327–329.

Rohe, D., Barrier, P., Clark, M., Cook, D., Vickers, K., & Decker, P. (2006). The benefits of pass-fail grading on stress, mood and group cohesion in medical students. *Mayo Clinic Proceedings, 81*(11), 1143–1148.

Schunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: from teaching to self-reflective practice*. New York: Guilford Press.

Scoles, P. (2008). Comprehensive review of the USMLE. *Advances in Physiology Education, 32*, 109–110.

Smith, K. A., Sheppard, S. D., Johnson, D. W., & Johnson, R. T. (2005). Pedagogies of engagement: Classroom based practices. *Journal of Engineering Education, 94*(1), 1–15.

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K., Kennon, M., & Deci, E. (2004). Motivating learning performance and persistance: The synergistic effects of intrinsic goal contents and autonomy-supported contexts. *Journal of Personality and Social Psychology, 87*(2), 246–260.

Vosti, K., & Jacobs, C. (1999). Outcome measurement in postgraduate year one of graduates from a medical school with a pass-fail grading system. *Academic Medicine, 74*, 547–549.

Weed, L. (1997). New connections between medical knowledge and patient care. *British Medical Journal, 315*, 231–235.

Weller, L. (1983). The grading nemesis: An historical overview and a current look at pass-fail grading. *Journal of Research and Development in Education, 17*(1), 39–45.

White, C. B. (2005). Self-regulated learning and self-assessment in medical education: Is it all Latin to medical students? Doctoral Dissertation, University of Michigan: AAT3186789.

White, C. B. (2007). Smoothing out transitions: How pedagody influences medical students' achievement of self-regulated learning goals. *Advances in Health Sciences Education, 12*(3), 279–297.

White, C. B., & Gruppen, L. D. (2007.) Self-regulated learning in medical education. Association for the study of medical education: Understanding medical education.

Zimmerman, B. J. (2000). Attaining self-regulation: Social-cognitive perspectives. In M. Boekaerts, P. R. Pintrich, & M. H. Zeidner (Eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press.

Zimmerman, B., Bonner, S., & Kovach, R. (1996). *Developing self-regulated learners: Beyond achievement to self-efficacy*. Washington, DC: American Psychological Association.