Impact of pass/fail grading on medical students' well-being and academic outcomes

Laura Spring,[1] Diana Robillard,[1] Lorrie Gehlbach[2] & Tiffany A Moore Simas[3,4]

OBJECTIVES Many medical schools are currently undergoing curriculum reform. When considering the means by which students will be evaluated in a revised curriculum, the need to reduce the prevalences of depression and anxiety associated with academic stress must be weighed against the importance of academic outcomes. Pass/fail evaluation, as compared with tiered grading, is commonly presented as a means to adequately assess student performance while minimising stress and anxiety. The purpose of this literature review was to determine the impact of pass/fail grading on medical student well-being and academic outcomes.

METHODS A systematic search was performed of the available literature published between January 1980 and August 2010, using the PubMed, Ovid Medline, Ovid PsycINFO and ERIC databases. Eligible papers assessed the impact of pass/fail grading on medical student well-being, academic outcomes or both. Academic outcomes included but were not limited to objective measures, such as performance on the US Medical Licensing Examination, and subjective measures, such as student desirability by residency programmes. Reference lists in identified papers were searched and all identified papers were run through a citation index.

RESULTS Four papers met the inclusion criteria for both well-being and academic outcomes. An additional five papers met the inclusion criteria for academic outcomes only. The four papers that focused on well-being reported improvement in specified areas. No significant difference was identified in any of the five papers examining objective academic outcomes or in those papers that examined the quality of residency programmes attained. Results from two studies suggested that some programme directors believe pass/fail grading creates disadvantages for students in attaining a residency, whereas a third study yielded mixed results about its impact on residency attainment.

CONCLUSIONS Student well-being is enhanced and objective academic performance is not adversely affected by a pass/fail evaluation system, but students' ability to obtain a desired residency programme may be hindered by individual programme directors' preferences for tiered grading systems. There is an overall paucity of literature on this topic and additional study is needed.

*Medical Education 2011: **45**: 867–877*
doi:10.1111/j.1365-2923.2011.03989.x

[1]University of Massachusetts Medical School, Worcester, Massachusetts, USA
[2]Center for Academic Achievement, Family Medicine and Community Health, University of Massachusetts Medical School, Worcester, Massachusetts, USA
[3]Department of Obstetrics and Gynaecology, University of Massachusetts Medical School, Worcester, Massachusetts, USA
[4]Department of Paediatrics, University of Massachusetts Medical School, Worcester, Massachusetts, USA

Correspondence: Tiffany A Moore Simas, Department of Obstetrics and Gynaecology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. Tel: 00 1 508 334 6678; Fax: 00 1 508 234 7774; E-mail: TiffanyA.MooreSimas@UMassMemorial.org

INTRODUCTION

Medical schools utilise a variety of grading systems. Referencing Association of American Medical Colleges (AAMC) data on US medical schools, Bloodgood *et al*.[1] recently reported that 40 schools use two grading intervals (usually pass/fail), 35 schools use three intervals (usually pass/fail/honours), 32 schools use four intervals (usually pass/fail/honours/high honours), and 26 schools use five intervals (usually A, B, C, D, F) during the pre-clinical years.

Numerous US and international medical schools have recently or are currently undergoing major curriculum reform and several new medical schools are opening in the USA. The impetus for curriculum reform stems from multiple aetiologies including but not limited to: (i) a call for undergraduate medical programmes to incorporate a vast array of new topics into traditional yet ever-expanding curricula that desire to be up to date and at the cutting edge, and (ii) societal concerns about health care disparities, especially with regard to ageing populations and end-of-life care.[2]

Curriculum design and reform are often discussed in association with grading systems and the quandary of how to most effectively evaluate students. The pass/fail grading system is often considered as one of the options for evaluation in the pre-clinical years, the clinical years or both. Newer pedagogy concurs with the concept of doctors as lifelong learners and emphasises the importance of self-directed learning.[3] An integral part of students' self-regulated learning is motivation.[4] Intrinsic motivation has been described as the intent to engage in learning on the basis of true interest and enjoyment, whereas extrinsic motivation is based on external rewards, such as grades.[4,5] In medical school, discriminating grades are the foundation of external motivation.[3] This is probably because grades and associated honour society induction are known to play key roles in residency selection criteria.[6–13] However, achieving high grades does not necessarily prepare students for the self-regulated learning they must undertake as practising doctors.

The pass/fail grading system for medical schools was first explored as an alternative to more traditional grading systems, such as A/B/C/D/F, grade point average (GPA) and class rank in the late 1960s as concerns arose that students were performing for 'the grade' rather than learning to improve knowledge.[14] Research revealed a lack of correlation between tiered grading and later clinical performance, thus supporting the use of pass/fail grading.[15,16] There was also evidence that pass/fail grading reduced competition and external motivation for grades without decreasing the amount of time students spent studying.[17] Over time, the trend has reverted to multi-tiered grading systems, particularly in the clinical years, as a result of insufficient discrimination between passing students.[14]

The benefits of pass/fail grading may include reduced stress, enhanced well-being, a less competitive learning environment and a greater focus on learning rather than on studying minutia purely for higher-grade achievement. Student well-being is not uniformly defined in the literature. For the purposes of this review, it is described as any aspect of physical, emotional, mental or spiritual health. Such benefits should not be underestimated as medical school-associated stress and increased rates of depression, as compared with rates in the general population, have been well documented in the literature.[18–24] Student distress in medical school may lead to feelings of stress, anxiety, burnout and depression, which can result in professional consequences such as impaired academic performance, academic dishonesty, declining empathy and medical errors. These feelings can also result in personal consequences such as broken relationships, poor self-care, substance abuse and suicide.[25] According to White and Fantone,[3] pass/fail grading has the potential to 'level the playing field' for incoming medical students with a variety of academic backgrounds and encourage collaboration and intrinsic motivation, both of which are keys to lifelong learning.

Concerns about the effectiveness of a pass/fail grading system may refer to a decline in class attendance, a decline in academic performance or effort, a decline in US Medical Licensing Examination (USMLE) scores, and less success in acquiring desired residencies.[1] Some programme directors may prefer to use more measurable, quantitative parameters to distinguish among students, and the lack of multi-tiered grading systems has the potential to raise the stakes of the USMLE Step 1 examination. Another concern is that pass/fail evaluation systems do not explicitly recognise excellence.[14]

There is limited literature on the impact of pass/fail grading in medical education; thus medical schools cannot easily make informed decisions regarding the implementation of or changing of grading systems. The existing literature examining the impact of pass/fail grading can broadly be separated into two

categories according to whether it refers to an impact on student well-being or an impact on academic outcomes. It is imperative to look at both aspects as short-term improved well-being in the pre-clinical years may conceivably be off-set by long-term poor academic outcomes. Our goal was to review all available literature with regard to pass/fail grading in medical education. Academic outcomes were considered broadly, and included impact on course grades, board scores, residency placement and success in residency.

METHODS

The databases PubMed, Ovid Medline, Ovid PsycINFO and ERIC (Educational Resources Information Centre) were searched. Searches were limited to original research articles published in English between January 1980 and August 2010. Initial searches were performed in each database utilising the following terms: 'medical education', 'medical', 'undergraduate medical', 'educational measurement', 'evaluation', 'evaluation nomenclature', 'evaluation system', 'grading', 'grading system', 'internship', 'residency', 'educational models', 'medical student', 'student assessment' and 'student evaluation'. To further search for papers related to well-being, the following search terms were also used: 'alcohol', 'anxiety', 'attitude', 'well-being', 'confidence', 'depression', 'distress', 'divorce', 'exercise', 'happiness', 'interpersonal', 'interpersonal relationships', 'psychological distress', 'mood', 'personal satisfaction', 'psychological', 'relationships', 'satisfaction', 'self efficacy', 'sleep', 'stress', 'substance', 'substance use' and 'wellness'. All search results were further limited by requiring the presence of the words 'pass' and 'fail'. The search of ERIC was limited to higher education articles only. To further identify papers related to academic outcomes, the following terms were searched in conjunction with 'internship' or 'residency' in PubMed and Ovid Medline: 'selection criteria', 'applicant selection' and 'applicant evaluation'. The six separate searches yielded a total of 759 papers, many of which were duplicates and were found in more than one of the searches.

Abstracts of all papers identified through the search criteria were reviewed by two of the authors (DR and LS). In instances in which the relevance of a paper could not be determined solely by reading the abstract or when no abstract was available, the full-text paper was reviewed. Any paper that did not specifically study the impact of pass/fail grading on either medical student well-being or medical student academic outcomes was excluded. Both well-being and academic outcome were broadly defined as indicated by the search terms. Academic outcomes considered for this review were anything that impacted the students' short-term or long-term success, including test scores in medical school and residency placement. Using the search terms and the inclusion and exclusion criteria described, we identified a total of four papers relating to well-being and eight papers relating to academic outcomes that were ultimately included in the final review. Four papers explored both topics. The references of selected papers were also checked for relevant papers that were not found in the original search. This yielded one additional paper relating to academic outcomes, which was included in the final review, bringing the number of papers on this topic to nine. All selected papers were then scrutinised by the Science Citation Index to see if other relevant papers had cited them. This did not yield any additional relevant papers.

RESULTS

Well-being

Four papers that met the inclusion criteria and specifically addressed issues of well-being in the context of pass/fail grading were identified through the search strategy (Table 1).[1,3,26,27] Each of the four papers found improvement in at least some of the well-being measures studied within the context of pass/fail grading. Definitions of well-being varied greatly, and referred to stress, anxiety and depression, as well as other concepts, such as self-control, good health, level of satisfaction, group cohesion and amount of time available for other activities (Table 1). Two studies found improvement in student satisfaction.[1,26] Student stress and anxiety were initially decreased with pass/fail grading according to Bloodgood *et al.*[1] However, this decrease did not persist and, by the fourth semester, levels of stress and anxiety in pass/fail graded students were not significantly different from those in a letter-graded cohort. By contrast, Rohe *et al.*[27] found that students who had experienced pass/fail grading in Year 1 continued to perceive less stress and reported greater cohesion even after entering Year 2, despite having returned to a five-tier grading system. Discrepant results were also found with regard to student extracurricular participation and pass/fail grading. Specifically, results from White and Fantone[3] revealed increases in students' time available for activities outside the classroom, including volunteer

Table 1 Studies investigating the impact of pass/fail grading on medical student well-being

Study	Participants and study design	Measures of well-being	Results
Bloodgood *et al.* (2009)[1]	Survey administered four times each to the class of 2006 (five-tier grading, *n* = 141) and to the class of 2007 (pass/fail grading, *n* = 140) during their first 2 years of medical school at the University of Virginia School of Medicine Response rate varied across administration times, ranging from 44% at the end of semester 3 (*n* = 62) to 75% at the end of semester 2 (*n* = 106) for the graded class, and from 69% at the end of semester 2 (*n* = 96) to 74% at the end of semester 1 (*n* = 103) for the pass/fail class	Web survey including the Dupuy[35] General Well-Being Schedule, which includes six subscales for anxiety, depression, positive well-being, self-control, vitality and general health Self-report survey items to measure class attendance and satisfaction with medical school as well as personal life	Semester 1: pass/fail class self-ratings indicated better well-being on all six subscales compared with students in a five-tier system Semesters 2 and 3: pass/fail class self-ratings indicated decreased anxiety and depression and an increase in positive well-being and vitality compared with students in a five-tier system Semester 4: no difference between groups Education satisfaction ratings were significantly higher in the pass/fail class No significant differences in self-reported class attendance, participation in enrichment activities, exercise or leisure activities
Robins *et al.* (1995)[26]	In 1992–1993 the University of Michigan Medical School revised its Year 1 curriculum and replaced its tiered grading system with pass/fail grading; it also instituted weekly quizzes Cohort 1: four-tier grading, *n* = 222 Cohort 2: pass/fail, *n* = 195 197 (89%) students in cohort 1 and 184 (94%) students in cohort 2 completed the satisfaction survey 167 (86%) of cohort 2 answered an open-ended question about what they liked most about their first year	Student satisfaction: students rated their level of satisfaction with the evaluation system and examination system, as well as the learning environment, on a 5-point Likert scale In addition, an open-ended item asked the pass/fail students what they liked best about their first year	The pass/fail cohort rated the evaluation and examination system more favourably than the four-tier grading cohort The pass/fail cohort rated the learning environment more highly 21% of respondents to the open-ended question wrote that the evaluation system was what they liked best about Year 1
White & Fantone (2009)[3]	At the University of Michigan, pass/fail grading was instituted in Year 2 in 2005–2006; previously pass/fail grading had applied only in Year 1 Members of the 2005–2006 Year 2 class were surveyed at the end of the school year Response rate was not reported	Participants rated a series of items designed to explore whether pass/fail grading freed up study time and how they used that time Items were rated on a 4-point scale (4 = strongly agree, 1 = strongly disagree)	Survey results revealed agreement ratings of 3.07 for 'more time to explore additional academic talents' (SD 0.75), 3.19 for 'more time to participate in volunteer/service activities' (SD 0.66), 3.21 for 'more time to participate in student organisations' (SD 0.70), 3.27 for 'more time to spend with family' (SD 0.69), and 3.43 for 'more time to exercise and improve personal wellness' (SD 0.59)

Table 1 (Continued)

Study	Participants and study design	Measures of well-being	Results
Rohe *et al.* (2006)[27]	Comparison of the class of 2005 (five-tier grading in Years 1 and 2, *n* = 41) with the class of 2006 (pass/fail system in Year 1 and five-tier grading in Year 2, *n* = 40) at the end of Years 1 and 2 At end of Year 1, 41 students (100%) in the class of 2005 and 39 (98%) in the class of 2006 participated At end of Year 2, 37 students (90%) in the class of 2005 and 38 (95%) in the class of 2006 participated	Perceived Stress Scale (10 items, Likert scale) Profile of Mood States (65 items, Likert scale, designed to measure six affective states in order to produce a total mood disturbance score) Perceived Cohesion Scale (six items, Likert scale, designed to measure sense of belonging and morale associated with group membership) Test Attitude Inventory (20 items, Likert scale, designed to measure test anxiety)	At the end of Year 1, students in the pass/fail system perceived less stress, had better overall mood and greater group cohesion compared with graded students There was no difference between the two groups in test anxiety at this time-point At the end of Year 2, students graded with pass/fail during Year 1 continued to perceive less stress and greater group cohesion There was no difference between the pass/fail and five-tier grading cohorts with regard to mood or test anxiety at the end of Year 2

SD = standard deviation

activities and personal time, whereas Bloodgood *et al.*[1] did not find an association.

Academic outcomes

Nine papers that met the inclusion criteria and specifically addressed issues of academic outcome in the context of pass/fail grading were identified through the search strategy (Table 2).[1,3,10,14,26–30] Medical school academic outcomes can be defined as specific objective measures of school performance, residency programme perceptions of student competitiveness or desirability, students' success in obtaining a desired residency and students' perceived performance in residency. We have dichotomised these academic outcomes broadly into 'academic achievement', such as GPA, class averages or scores on the USMLE, and 'residency attainment and performance'.

Academic achievement

Bloodgood *et al.*,[1] comparing a cohort assessed using a pass/fail system with a graded cohort, found no difference in any of the academic outcomes measured, including pre-clinical GPA, clerkship grades or USMLE Step 1 and 2 scores. Robins *et al.*[26] reported no significant difference in the anatomy grades (the only course that allowed direct comparison) of pass/fail and tiered-grading cohorts. No direct comparison between the cohorts was possible for grades achieved in other courses (the courses were significantly modified as part of curricular revision), but the pass/fail cohort achieved averaged numeric grades that were significantly higher than the pass/fail cut-off of 75%, alleviating concerns that pass/fail grading would lead students to choose not to strive for excellence.[26] White and Fantone[3] found that two Year 2 end-of-course averages declined and one improved when the grading system was switched from an honours/high pass/pass/fail system to a pass/fail format; there was no significant difference in USMLE scores. Rohe *et al.*[27] also reported no significant difference in USMLE scores.

Residency attainment and performance

Tardiff[30] found that for the areas of 'medical knowledge and skill', 'initiative and motivation' and 'personal relations with patients', 76%, 77% and 91% of programme directors, respectively, from various fields believed that pass/fail students performed at the same level as tier-graded students. According to Vosti and Jacobs,[28] when residency directors were asked to compare the clinical preparedness of residents graduating from medical schools with

Table 2 Studies investigating the impact of pass/fail grading on medical student academic outcomes

Study	Participants and study design	Academic outcome measurements	Results
Bloodgood *et al.* (2009)[1]	Objective data were collected on various academic outcomes in the class of 2006 (five-tier grading system, *n* = 141) and the class of 2007 (pass/fail grading, *n* = 140) at the University of Virginia School of Medicine	Performance in basic science courses Clerkship grades USMLE Step 1 and Step 2 CK Quality of the residency programme to which the student was matched Quality measured by percentage of residents in the programme passing the board certification examination Data only available for IM, FM, paediatrics and general surgery	Overall basic science average did not differ between the two classes No significant difference in clerkship grades No significant difference on Step 1 or Step 2 CK No significant difference between pass/fail and tiered grading groups in the quality of residency programmes obtained
Vosti & Jacobs (1999)[28]	Residency PDs rated graduates of Stanford compared with their peer residents on their clinical preparedness to enter residency Directors of the postgraduate year 1 for all 169 graduates from the Stanford classes of 1993 and 1994 who sought further clinical training were asked to participate 144 (85%) of the residency PDs participated Stanford employs a pass/fail system, without honours or class rankings, for rating the performance of students in both the basic and clinical sciences	Using a rating scale of 1–9 (9 is highest), six cognitive factors (history taking, physical examination, performance of procedures, medical knowledge, clinical judgement, application of knowledge) and four non-cognitive factors (interpersonal relations with professionals, patients and peers; dependability) were rated	Compared with their peer group's clinical preparedness, 3% of Stanford graduates were rated as 'poor', 20% as 'good', 44% as 'excellent' and 33% as 'outstanding' Of note, stratification of programmes by either hospital or medical specialty did not reveal significant differences in overall clinical competence
Robins *et al.* (1995)[26]	In 1992–1993 the University of Michigan Medical School revised its Year 1 curriculum and replaced its tiered grading system with a pass/fail format Cohort 1: four-tier grading, *n* = 222 Cohort 2: pass/fail, *n* = 195 All student data were used for course performance analyses	Content mastery was determined through course averages (percentage correct in multiple-choice items) Only one course, anatomy, provided exact comparison data because all other courses were changed under the curriculum revision	For the pass/fail cohort, all final course averages were significantly higher than the pass cut-off of 75% No significant difference found in anatomy course final average between the pass/fail and tiered grading cohorts
White & Fantone (2009)[3]	At the University of Michigan, a class using a multi-tiered grading scale in Year 2 (2004–2005) was compared with a class using a pass/fail grading scale in the same curriculum in Year 2 (2005–2006)	Performance on Year 2 examinations, USMLE Steps 1 and 2, and residency placement Quality of residency programme was determined by sending NRMP results for the 2004–2005 Year 2 class and the 2005–2006 Year 2 class to each of the residency PDs at the University of Michigan in order to rank each programme in its specialty	In 11 Year 2 classes, the pass/fail cohort performed significantly lower in two and significantly higher in one of the courses There was no statistical difference between the two classes in residency placement There was also no statistically significant difference in how many students from each class ranked into a top 15 residency programme

Table 2 (Continued)

Study	Participants and study design	Academic outcome measurements	Results
Dietrick *et al.* (1991)[14]	All 302 US general surgery PDs listed in the 1989 'Green Book' were invited to complete the survey Response rate was 64%	Survey designed to measure whether the evaluation system used by a medical school (pass/fail versus competitive grading) had an effect on a student's ability to compete for a general surgery residency	81% of PDs believed that the medical student's ability to compete for a residency position was adversely influenced by the pass/fail method of evaluation
Provan & Cuttress (1995)[29]	Questionnaire sent to all 110 PDs of residency training programmes in Ontario in June 1993 Of the 110 PDs, 96 (87%) responded 92 (84%) responded to the questions regarding grading preferences	Regarding pass/fail grading, one question asked which type of grading system was preferred and another question asked whether or not students from schools with pass/fail grading systems would be disadvantaged compared with those from schools with letter-based or numeric grading systems	35 (38%) preferred a numeric grading system, 26 (28%) a letter-based grading system, 23 (25%) an honours/pass/fail system and eight (9%) a pass/fail system 61 (66%) felt that students applying to their programme from a school that used a pass/fail system would be at a disadvantage compared with students from schools that used a letter-based or numeric grading system
Rohe *et al.* (2006)[27]	Compared the class of 2005 (five-interval system, $n = 41$) and the class of 2006 (pass/fail system in Year 1 and five-interval system in Year 2, $n = 40$)	USMLE Step 1 scores	Scores did not differ statistically significantly between the two groups (five-tier grading system mean: 234 ± 22; pass/fail mean: 227 ± 21)
Hughes *et al.* (1983)[10]	879 applicants for IM residency programme at McGaw Medical Center, Northwestern University, who were ranked for NRMP during four consecutive cycles in 1978–1981 An additional 184 applicants (17%) who were ranked for the NRMP during the same period were not included in the analysis because of missing items 102 applicants (11%) came from pass/fail grading systems, 630 (72%) from honours/pass/fail grading systems and 147 (17%) from letter-graded systems	A composite faculty rating for each applicant was generated using a 10-point scale based upon overall grade, Year 3 medicine clerkship grade, AOA, NBME Part 1, LORs, MSPE and interview in order to create the NRMP list Using the *F*-test, grading system was used as an independent variable and faculty rating and NRMP rank as dependent variables	The type of grading system was not significantly associated with the faculty rating or NRMP rank of applicants

Table 2 (Continued)

Study	Participants and study design	Academic outcome measurements	Results
Tardiff (1980)[30]	Random sample of 940 residency programmes, stratified by geographic region and specialty, were selected from the *Directory of Residency Training Programs* (1978–1979) and were sent a questionnaire in 1979 Response rate was 81% and 760 responses were included in the analytic subset Specialties were FM, IM, Ob/Gyn, paediatrics, psychiatry and general surgery	Relevant questions were: 'In selecting residents, do you give preference to students from graded or pass/fail schools?' 'Concerning the performance of residents in your programme, do those from pass/fail schools perform at a higher, lower or same level as those from graded schools in the areas of medical knowledge and skills, initiative, and motivation, and interpersonal relations with patients?'	For all specialties combined, 73% of PDs indicated that they do not give preference to students from either graded or pass/fail schools when selecting residents 27% indicated a preference for students from schools using tiered grading systems Programmes in medicine, Ob/Gyn and surgery were more likely than those in the other specialties to prefer students from schools using tiered grading systems Of the programmes that filled all their Year 1 positions (66%), 33% were more likely to prefer students from schools using tiered grading systems Regarding medical knowledge and skill, initiative and motivation, and personal relations with patients, 76%, 77% and 91%, respectively, of PDs felt that pass/fail students performed at the same level as their peers from schools utilising tiered grading PDs who rated the medical knowledge and skills of pass/fail students lower than those of graded students also indicated that they gave preference to students from schools using tiered grading systems

AOA = Alpha Omega Alpha (medical honour society); CK = clinical knowledge; FM = family medicine; IM = internal medicine; LORs = Letters of Recommendation; MSPE = medical student performance evaluation; NBME = National Board of Medical Examiners; NRMP = National Resident Matching Programme; Ob/Gyn = obstetrics and gynaecology; PD = programme director; USMLE = US Medical Licensing Examination

pass/fail systems with that of their peers graduating from schools with tier-grading systems, 3% were rated as 'poor', 20% as 'good', 44% as 'excellent' and 33% as 'outstanding'. With regard to obtaining residency positions, Tardiff[30] found that 73% of directors reported that they did not give preference to either students from schools using tier grading or students from schools using pass/fail systems, but 27% indicated that they preferred to accept students from tier-grading schools. Tardiff[30] also reported that 33% of programmes that filled all their first-year spots (hence those for which entry is likely to be more competitive) preferred to accept students from tier-grading schools. Bloodgood *et al.*[1] noted no significant difference between the two cohorts in the quality of residency programmes obtained. White and Fantone[3] also found no statistical difference between pass/fail and tier-graded cohorts in residency placement, nor in how many students from each cohort ranked into a top 15 residency programme. In a survey of programme directors in Ontario, Provan and Cuttress[29] found that 66% felt that students applying to their

programme from a school that used a pass/fail system would be at a disadvantage. Dietrick *et al.*[14] reported that 81% of general surgery programme directors believed that ability to compete for a residency position was adversely influenced by the pass/fail method of evaluation.

DISCUSSION

The goal of this study was to review all the available literature examining the impact of pass/fail grading in medical school on student well-being and academic outcomes. We considered it imperative to look at both aspects because short-term improved well-being in the pre-clinical years may conceivably be off-set by long-term poor academic outcomes. Our literature review suggests that student well-being is enhanced and objective academic performance is not adversely affected by a pass/fail evaluation system. However, students' ability to obtain a place in a desired residency programme may be hindered by individual programme directors' preferences for tiered grading systems.

Student stress and anxiety are reduced, at least initially, by pass/fail grading, but long-term data are lacking. Bloodgood *et al.*[1] found that by the end of semester 4 (end of Year 2), levels of stress and anxiety were not significantly different from those in graded cohorts, which directly contrasts with results reported by Rohe *et al.*,[27] who found that these benefits persisted, even once students had resumed a five-tier grading system in Year 2 of medical school. The end of semester 4 traditionally coincides with the administration of USMLE Step 1, which is a major confounding factor when trying to determine the source of students' perceived stress. Additionally, the study by Bloodgood *et al.*[1] included cumulative academic honours for 20% of the class at the end of Year 2, which may also explain the lack of reduced stress and anxiety at the end of semester 4. The inherent weakness in the review of well-being is the wide variety of ways in which each study examined and defined student well-being. No two papers used the same tools to measure well-being. More systematic and comparable studies are needed in order to facilitate the drawing of stronger conclusions.

In the 1970s, Moss *et al.*[31] brought the pass/fail debate into the residency selection discussion when they reported that surgery residents from medical schools using pass/fail systems performed inferiorly to those from medical schools using graded formats. This study is over 30 years old and examined only one programme. This review demonstrates that no academic outcome study since 1980 has provided longitudinal objective data demonstrating the effect of pass/fail grading on students' ability to obtain a place at a desired residency or to perform in that residency. Tardiff[30] noted that, although only 27% of programme directors gave preference to students from tier-grading schools, the directors of more desirable programmes were more likely to prefer such students. These data were opinion-based and no retrospective review of applicants to desirable residencies was completed to determine whether this self-reported preference was actually acted upon during the resident selection process. It is possible that more competitive residency programmes may place greater emphasis on grades; this is a potential area for further research.

Variation in the way that pass/fail grading was defined across studies significantly impacts the conclusions that can be drawn on associations between grading systems and academic outcomes. Of the reports that studied a single school, some examined the impact of pass/fail grading in particular years in medical school and others studied schools with pass/fail grading across all years of the curriculum. Furthermore, many of the data on performance in residency are based on self-reporting by residency directors rather than direct measurement of clinical performance. The questions posed to programme directors were often broad and, again, did not typically distinguish whether or not the residents came from schools that used pass/fail grading in select years or in all years. Programme directors were not blind to grading system status and therefore it is possible that being aware that a resident came from a school that used a pass/fail system led the director to give lower ratings. Blinded studies, perhaps of attending doctors who are not familiar with the grading practices of their residents' medical schools, are needed to further clarify this issue.

Given the concern that pass/fail grading does not clearly recognise or encourage excellence, Peters and Finch[32] examined ways in which outstanding performance is acknowledged by surveying medical schools with non-traditional grading systems. It was found that most schools with non-traditional grading systems rely on institutional awards, including clerkship honours and election to honour societies such as Alpha Omega Alpha (AOA). A confounding factor in several of the studies is whether the pass/fail system adopted by a particular school is truly transparent. For example, if a school uses a pass/fail system in the pre-clinical years, but students' scores are still used to determine their receipt of honour society induction, institutional awards or external

grants and scholarships, then the benefits of pass/fail grading may conceivably be lost. This may also be true in schools, such as the University of Virginia School of Medicine, in which students in the pass/fail cohort are still ranked so that the top 20% of students, based on class scores averaged across Years 1 and 2, are awarded cumulative honours. Of those students who reported making a conscious effort to pursue cumulative honours, 70% said that the decision increased stress, whereas of those who reported making a conscious decision not to pursue honours, 92% reported reduced stress.[1] It is noteworthy that the school, with the support of students, ultimately chose to drop the cumulative honours system.[1]

The purposes of this review were to determine if pass/fail grading systems benefit students in the short-term by positively affecting their well-being and to evaluate their effects on long-term academic outcomes. It is not surprising that this review demonstrated improved well-being in students who experienced pass/fail grading. A balance between improved well-being and maximising desired academic outcomes must be found. This review demonstrates that, overall, pass/fail grading improves student well-being and does not compromise academic outcomes.

It has been argued that the decreased stress and competition achieved by a pass/fail grading system may lead to efforts to seek alternative means of distinction, such as by means of extracurricular activities or research endeavours.[33] Although this approach might be considered to encourage more well-rounded students, the shifting of stress and competition from one environment to another may jeopardise the benefits of pass/fail grading in terms of well-being.[33] As far as we are aware, this has not been directly studied and may be an area for future research. A compromise appears to be a hybrid system, in which pass/fail grading is applied in some years or portions of medical school and systems with more than two grading intervals are applied in others. Several schools already use such a model. A common approach involves using a pass/fail system in the pre-clinical years and a more discriminating grading scale in the clinical years. An alternative model also employed by several schools involves using a pass/fail system in Year 1 only, which can help to 'level the playing field', as noted by White,[3] thus allowing former biochemistry and music majors equal opportunities at classroom success.

As previously discussed, issues of transparency regarding honour societies and other cumulative honours and awards need to be considered. Given that the USMLE Step 1 is an objective, familiar measurement, it can serve as a surrogate for pre-clinical grades in the residency selection process. If programme directors were to use it as such, students from pass/fail grading systems may not be disadvantaged in comparison with their graded counterparts.[6] This, however, maintains the USMLE Step 1 as an extremely high-stakes examination. Of note, a survey by Wagoner *et al.*[8] of 275 programme directors found that 56% thought that honours grades in pre-clinical courses were more important than the USMLE Step 1 examination. However, these data were sourced from a period several years ago and represent the opinions of a relatively small subset of programme directors. The 2010 National Resident Matching Programme survey of programme directors revealed that, for all specialties combined, the mean importance of USMLE Step 1 and honours grades in pre-clinical courses were rated as 4.1 and 3.1, respectively, on a scale of 1–5, where 1 = not at all important and 5 = very important.[34]

This review, although it is comprehensive and based on existing literature, reveals limited resources regarding the impact of pass/fail grading on student well-being and academic outcomes. This paucity is further highlighted by the dearth of recently published literature and thus this review may not accurately reflect current medical school curricula in the wake of recent reforms. Additionally, generational changes in the views of both educators and students on this issue may not be well represented in the available literature. Given the considerable curriculum reform and associated changes in student evaluation taking place in the USA and elsewhere, now is an important and opportune time to study the effects of different grading systems. More data will help to inform choices made by medical schools regarding grading practices. In addition, as illustrated by this review, an effective, reproducible means of studying both short- and long-term well-being in medical students is needed.

Contributors: LS and DR contributed to the devising of the study, data acquisition and interpretation, and the drafting of the manuscript. LG and TAMS contributed to the devising of the study, data interpretation and the critical revision of the manuscript. All authors approved the final manuscript for publication.
Acknowledgements: none.
Funding: none.
Conflicts of interest: none.
Ethical approval: not applicable.

REFERENCES

1 Bloodgood RA, Short JG, Jackson JM, Martindale JR. A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Acad Med* 2009;**84** (5):655–62.
2 Anderson MB, Kanter SL. Medical education in the United States and Canada, 2010. *Acad Med* 2010;**85** (Suppl 9):2–18.
3 White CB, Fantone JC. Pass-fail grading: laying the foundation for self-regulated learning. *Adv Health Sci Educ Theory Pract* 2010;**15** (4):469–77.
4 White CB. Smoothing out transitions: how pedagogy influences medical students' achievement of self-regulated learning goals. *Adv Health Sci Educ Theory Pract* 2007;**12** (3):279–97.
5 Vansteenkiste M, Simons J, Lens W, Sheldon KM, Deci EL. Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *J Pers Soc Psychol* 2004;**87** (2):246–60.
6 Green M, Jones P, Thomas JX Jr. . Selection criteria for residency: results of a national programme directors survey. *Acad Med* 2009;**84** (3):362–7.
7 Wagoner NE, Suriano JR. Programme directors' responses to a survey on variables used to select residents in a time of change. *Acad Med* 1999;**74** (1):51–8.
8 Wagoner NE, Suriano JR, Stoner JA. Factors used by programme directors to select residents. *J Med Educ* 1986;**61** (1):10–21.
9 Featherstone HJ, Ramsey PG. Analysis of selection criteria for medical residents. Differences between primary care and traditional pathway committees. *Am J Med* 1983;**75** (4):687–90.
10 Hughes RL, Golmon ME, Patterson R. The grading system as a factor in the selection of residents. *J Med Educ* 1983;**58** (6):479–81.
11 Taylor CA, Weinstein L, Mayhew HE. The process of resident selection: a view from the residency director's desk. *Obstet Gynecol* 1995;**85** (2):299–303.
12 Bajaj G, Carmichael KD. What attributes are necessary to be selected for an orthopaedic surgery residency position: perceptions of faculty and residents. *South Med J* 2004;**97** (12):1179–85.
13 Otero HJ, Erturk SM, Ondategui-Parra S, Ros PR. Key criteria for selection of radiology residents: results of a national survey. *Acad Radiol* 2006;**13** (9):1155–64.
14 Dietrick JA, Weaver MT, Merrick HW. Pass/fail grading: a disadvantage for students applying for residency. *Am J Surg* 1991;**162** (1):63–6.
15 Wingard JR, Williamson JW. Grades as predictors of physicians' career performance: an evaluative literature review. *J Med Educ* 1973;**48** (4):311–22.
16 Korman M, Stubblefield RL. Medical school evaluation and internship performance. *J Med Educ* 1971;**46** (8):670–3.
17 Jessee WF, Simon HJ. Time utilisation by medical students on a pass-fail evaluation system. *J Med Educ* 1971;**46** (4):275–80.
18 Wolf TM. Stress, coping and health: enhancing well-being during medical school. *Med Educ* 1994;**28** (1):8–17; discussion 55–7.
19 Radcliffe C, Lester H. Perceived stress during undergraduate medical training: a qualitative study. *Med Educ* 2003;**37** (1):32–8.
20 Rosenthal TL, Edwards NB, Ackerman BJ. Student's self-ratings of subjective stress across 30 months of medical school. *Behav Res Ther* 1987;**25** (2):155–8.
21 Rosenthal TL, Rosenthal RH, Edwards NB. Students' self-ratings of stress in medical school: a replication across 20 months. *Behav Res Ther* 1990;**28** (2):171–3.
22 Rosal MC, Ockene IS, Ockene JK, Barrett SV, Ma Y, Hebert JR. A longitudinal study of students' depression at one medical school. *Acad Med* 1997;**72** (6):542–6.
23 Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among US and Canadian medical students. *Acad Med* 2006;**81** (4):354–73.
24 Goebert D, Thompson D, Takeshita J, Beach C, Bryson P, Ephgrave K, Kent A, Kunkel M, Schechter J, Tate J. Depressive symptoms in medical students and residents: a multi-school study. *Acad Med* 2009;**84** (2):236–41.
25 Dyrbye LN, Thomas MR, Shanafelt TD. Medical student distress: causes, consequences, and proposed solutions. *Mayo Clin Proc* 2005;**80** (12):1613–62.
26 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** (4):327–9.
27 Rohe DE, Barrier PA, Clark MM, 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** (11):1443–8.
28 Vosti KL, Jacobs CD. Outcome measurement in postgraduate year one of graduates from a medical school with a pass/fail grading system. *Acad Med* 1999;**74** (5):547–9.
29 Provan JL, Cuttress L. Preferences of programme directors for evaluation of candidates for postgraduate training. *CMAJ* 1995;**153** (7):919–23.
30 Tardiff K. The effect of pass-fail on the selection and performance of residents. *J Med Educ* 1980;**55** (8):656–61.
31 Moss TJ, Deland EC, Maloney JV Jr. . Selection of medical students for graduate training: pass/fail versus grades. *N Engl J Med* 1978;**299** (1):25–7.
32 Peters MJ, Finch WT. Recognition of outstanding performance in medical schools with non-traditional grading systems. *J Med Educ* 1984;**59**:514–6.
33 Guo R, Muir J. Toronto's honours/pass/fail debate: the minority position. *UTMJ* 2008;**85** (2):95–8.
34 National Resident Matching Program, Data Release and Research Committee. *Results of the 2010 NRMP Program Director Survey*. Washington, DC: NRMP 2010.
35 Dupuy HJ. The general well-being schedule. In: McDowell I, Newell C, eds. *Measuring Health: A Guide to Rating Scales and Questionnaire*, (2nd edn). USA: Oxford University Press 1977; 206–13.

Received 26 November 2010; editorial comments to authors 21 January 2011; accepted for publication 9 February 2011