An investigation of pre-clinical medical students' preference for summative or formative assessment for physiology learning.

Mark G. Rae[1#] and Mohammed H. Abdulla[1]

1. Department of Physiology, University College Cork, Western Gateway Building, Cork, Ireland.

author for correspondence

e-mail address: m.rae@ucc.ie

Running title: **Assessment type preference of pre-clinical medical students**

Key Words: summative, formative, pre-clinical, assessment, physiology

New and Noteworthy

Although several studies have investigated the relative effectiveness of regular summative *versus* formative assessments as tools to facilitate student learning and understanding, ours is the first to explicitly solicit students' views on, and preferences for, either regular formative or summative assessment in pre-clinical medicine. We demonstrate that, overall, students preferred formative- to summative-type assessments, due to the immediacy of feedback, but also that summative tests incentivized them to study and keep up with material more.

Abstract

An investigation of pre-clinical medical students' preference for summative or formative assessment for physiology learning.

Both summative and formative assessments are known to facilitate student learning and understanding and helps them to identify areas of weakness. However, few studies have investigated students' preference for either summative or formative evaluations, particularly in the area of pre-clinical medicine. The current study addresses this deficit by surveying 137 first year graduate entry to medicine (GEM) pre-clinical medical students from two consecutive years (2018-19 & 2019-20), for their thoughts on the six summative (*i.e.* for a small percentage of marks), proctored, and the fiver informal, formative (*i.e.* no marks available) continuous assessments in physiology that they encountered in semesters 1 and 2, respectively.

Our survey revealed that between 75-90% of students felt that both evaluation formats were roughly equally useful (*i.e.* selecting options, 'Agree' or 'Strongly Agree') both for providing feedback about their understanding of physiology, and for identifying deficits in their physiology knowledge. However, although a significantly larger number of students felt that summative evaluations motivated them to study more than the formative evaluations (p=0.006), overall, more students favored formative over summative assessments. Notably however, GEM students from non-biomedical backgrounds were significantly more in favor of summative assessments than those from either biomedical backgrounds (p=0.003) or the whole GEM survey cohort (p=0.01).

The implications of these findings will be discussed, with suggestions as to how the student views outlined here might be facilitated within an academic program to maximise both student learning as well as their motivation to study and keep up with taught material.

Introduction

It has been documented previously that pre-clinical medical students generally find physiology to be the most challenging of the pre-clinical life sciences to learn (1). This is probably due in large part to the fact that physiology does not lend itself easily to simple rote memorization, and instead relies more on the understanding and application of physiological principles to problem-solving. As such, numerous strategies have been employed by educators to assist students with their understanding, acquisition and correct application of physiological concepts. However, the most commonly utilized strategy is to provide students with regular evaluations of their knowledge, usually in the form of either summative or formative assessments (2). Although precise definitions for the terms 'summative' and 'formative' vary depending on the context in which they are deployed (2, 3), with regard to the present study, summative assessments are defined as tests (taken individually, with no discussion of questions/answers with peers or educators), that take place in a proctored environment at the end of set periods/topics of teaching, for a certain percentage of module marks. Feedback for these types of tests, if provided at all, can often be generic and delivered days or weeks after the tests have taken place.

In contrast, formative assessments herein are defined as informal, no stakes, tests of knowledge that are primarily utilized to provide both students and educators with a means of gauging students' understanding of taught material, in which feedback (on an individual's or class's performance) is provided immediately or soon after a test has been taken (2).

Numerous studies have shown that regular assessments are beneficial to student learning and motivation (4-11) probably due, at least in part, to the testing effect (12).

In relation to the current study, several studies have inferred from survey data that students prefer formative to summative assessments (4, 13), with recognition of a clear correlation between formative assessments and student motivation and achievement (6, 14-16). However,

to the best of our knowledge, no previous investigations have either explicitly determined students' preference for formative or summative assessment, or determined the effect of utilising regular, or semi-regular, cumulative summative assessments on students' perceptions of preparedness for terminal high stakes end of semester examinations. Specifically within the context of pre-clinical medicine, which is the focus of the current study, this is somewhat surprising given that not only can student learning strategies and exam performance be affected by their assessment preference type (17-22), albeit that this is a relatively weak correlation (21) but also because medical education is centered upon regular student assessment (23, 24). As such, although students' assessment preferences should not be the sole, or even the main, guiding principle for faculty when making decisions about assessment structure for particular material/circumstances, due to the fact that a student's preferred choice may not necessarily be what best facilitates their learning and understanding, it is reasonable to propose that student perceptions should at least be considered in order to maximally facilitate student understanding and accurately reflect student learning/knowledge. Assessments should also provide motivation for students to study and keep up with taught material. This particularly applies to pre-clinical medical students, where factual understanding/knowledge is prioritized, as opposed to later clinical stages, where greater emphasis is placed on critical thinking skills (25, 26).

We have been aware for several years that the vast majority of graduate entry to medicine (GEM) students at UCC very much enjoyed the regular physiology assessments that are provided in their first year (27, 28). However, it was unclear if they had a particular preference for one type of regular assessment format over the other: summative, with limited, delayed feedback, or formative, with immediate feedback. Although we hypothesized that students would prefer the latter, the aim of our study was to determine if such a preference existed and, if so, if it was context-specific.

1

Methods

The current study was undertaken at University College Cork (UCC), Ireland, with data collected from two separate surveys conducted over the course of two academic years (2018-19 & 2019-20), with input from two separate first year GEM cohorts, which includes students of different ages (21-40), sex (male and female), country of origin (EU and non-EU) and degree backgrounds (subdivided here for simplicity into biomedical and non-biomedical). Ethical approval was obtained for each survey, separately, from University College Cork Social Research Ethics Committee (approvals 2018-130 & 2019-187). Students entering the GEM program at UCC must possess at least one undergraduate degree (from any discipline), at a minimum grade of a second class honors, grade one, or a 3.3 grade point average. All first year GEM students in both years that the survey was run were eligible to take part in the study.

Each survey was undertaken by GEM students following the first two modules of the pre-clinical GEM program, GM1001 (Fundamentals of Medicine I; 13 teaching weeks) and GM1002 (Fundamentals of Medicine II; 11 teaching weeks). The physiology curriculum taught within these modules consists of 13 lectures and 10 laboratory practical hours in GM1001 and 20 lectures and 14 laboratory practical hours in GM1002, with teaching divided roughly equally across both modules in 2018-19 and 2019-20, between two of the authors, MGR and MHA. Furthermore, the same lectures and continuous assessments (CAs) were delivered and conducted by the same two educators (MGR and MHA) across both years of the study.

Regular CAs consisting of up to 20 questions in a 'single best item' (SBI) format, where a question stem is answered by selecting one of up to five possible answers, were run throughout both GM1001 (summative) and GM1002 (formative), with six in GM1001 and five in GM1002. The SBI format was utilized not only because it is the same format as in their formal end of module and year exams and also because many studies have shown that this is amongst

the most popular formats of assessment both with pre-clinical (10, 25) and students in general (17, 29, 30) as students perceive them to be easier to study for, easier to take and will result in higher scores than more complex assessment types such as essay-type or constructed response-types of questions (10, 29).

Both summative and formative CAs were taken by the students in both year groups, on average, one to two weeks after in-person, didactic lectures on the relevant material had been delivered to the class. These lecture sessions were supplemented by supporting material posted on UCC's learning management system, Canvas, that included the lecture slides (PowerPoint and PDF + slide note versions), pre-records of the lectures, supplementary videos on lecture material and learning objectives to guide student review of the material.

The SBI questions used in the CAs for both GM1001 and GM1002 were a mixture of pure recall, comprehension- and application-type questions in line with Bloom's taxonomy (31) and consistent with the format and difficulty of questions used in their formal end of module, and year, examinations. However, the mode of delivery of the CAs differed between the two modules. Specifically, in GM1001 the 20-question CAs (accessed *via* a secure browser and restricted to members of the GEM 1 classes only, secured by logging in to a lab computer with their unique student identification number), for which 30 minutes were provided to complete each test, were conducted on individual computers by each student in the same proctored setting on each occasion. Upon submission of each CA, automated feedback provided students only with their score and which questions they answered correctly. Specific feedback on each question was provided by the lecturer in scheduled tutorials 1-3 weeks after each CA. Importantly, in the context of the current study, these evaluations were summative, in that the CAs constituted 1% of the students' end of module Physiology exam grade (*i.e*. each CA was worth 0.2%). All students took all of the GM1001 CAs.

In GM1002, CAs were conducted in the same teaching laboratory as the GM1001 CAs, were again restricted only to members of the GEM 1 classes, and accessed by logging in to a lab computer. However, the formative CAs were administered by means of the free, cloud-based student response system (SRS), Socrative® (28, 32-37). The primary advantages of using this type of SRS, aside from its ease of use and popularity with students (28), are that all students can anonymously answer each question at the same time, and that it also provides both students and faculty with immediate feedback on how the class has answered each question. For each of the five CAs in GM1002, students were again provided with 90 seconds to answer each question (as in the GM1001 CAs), but the time was either occasionally shortened, if all of the logged students had answered the question (Socrative provides a live display both of how many participants have logged into the quiz and also how many have answered the question being displayed) or, rarely, increased up to a maximum of 120 seconds if more than 10% of the class had still not provided an answer after 90 seconds. Although students were permitted to discuss questions with their peers (to encourage peer-assisted learning (38, 39)), and could log their answers either as individuals or as a group, only 32.6% of students participated in the formative CAs as part of group of two or more (*data not shown*), which, we believe, allows reasonable comparisons to be made between the two types of assessments. In contrast to the CAs conducted in GM1001, the GM1002 CAs were purely formative (*i.e.* test scores made no contribution to students' final physiology grade).

Due to the requirement to explain answers to the formative GM1002 CA questions, it generally took longer to cover each question than for the GM1001 computer-based assessments. As such, in the one-hour time slots provided for each CA in GM1002, a maximum of 15 questions (minimum 12 questions) were covered.

Data Collection and Analysis

The entire first year GEM class, consisting of 81 students in 2018-19 and 84 students in 2019-20, was provided with access to an itemised survey posted on Google docs after the final teaching session of the GM1002 module in March 2019 and March 2020, respectively, with the same survey used to assess both summative and formative assessments. All GEM students were encouraged to complete the survey, but participation was voluntary, with students assured that they could respond as freely as they wished as all data were completely anonymous. The survey remained available to the students for two months before the data were downloaded for analysis. The survey instrument consisted of a total of 42 items which, in addition to collecting demographic data (four items), sought information both on students' perceptions of the GM1001 (summative, with up to 1% of marks going towards their end of module Physiology exam grade) *versus* GM1002 (formative) CAs (15 items), as well as their views on the ethics of cheating on the summative assessments (19 items). Data from the second element on students' views on the ethics of cheating in small/no stakes CAs has not been included in the current manuscript. The 15 survey items pertaining to the physiology CAs focussed on determining if students felt that they were useful for their learning and understanding of physiology and, if so, if their opinion differed for summative *vs* formative. The survey was checked for clarity by a colleague and was tested by two graduate students, with no problems indicated.

For statistical comparisons, all data from Likert scale responses were converted to ordinal data, with "strongly agree" = 5, "agree" = 4, "neither agree nor disagree" = 3, "disagree" = 2, "strongly disagree" = 1. We also gathered demographic data such as student sex, age, and educational background (*i.e.* biomedical *vs* non-biomedical). All survey data from respondents were downloaded into Microsoft Excel for statistical analyses. Unless specifically stated otherwise, all statistical comparisons between groups for summative *vs* formative CAs were conducted using Student's unpaired t-test (assuming equal variance). Analysis of Likert data

for each survey item was conducted using the χ^2 -test to determine significant deviations from chance, with the numbers of students who selected options “Agree” and “Strongly agree” being pooled and analysed against pooled options “Disagree” and “Strongly Disagree”.

Data in the text are presented as mean (standard deviation), with values of $p < 0.05$ being considered statistically significant for both t- and χ^2 - tests.

Results

Student demographics

For the current study, 67/81 (82.7%) and 68/84 (80.9%) students from each first-year class completed, or partially completed, the survey in 2019 and 2020, respectively, with demographics broadly similar across both years. For example, in 2018-19, 34 respondents (50.7%) were female, with 33 (49.3%) male, and in 2019-20, 37 respondents (54.4%) were female and 31 (45.6%) male. Thirty two (47.8%) and 34 respondents (50%) indicated that their region of origin was Europe (EU) in 2018-19 and 2019-20, respectively, while 34 (50.7%) and 33 (48.5%) were from non-European (non-EU) countries, respectively (one unanswered)(Figure 1A). Furthermore, there were significantly greater numbers of students who regarded their prior degree backgrounds as biomedical rather than non-biomedical, with an overall distribution of 70.4% biomedical *vs* 29.6% non-biomedical (Figure 1A). The majority of students fell within the 22-25 age band (68.7%) (Figure 1B).

Students' attendance of GM1001 summative and GM1002 formative CAs

All students attended all five summative GM1001 physiology CAs as attendance was compulsory for both survey years (students who missed CAs for specific, verifiable reasons, were allowed to take them upon their return to campus). Although attendance at the five formative GM1002 CAs was not similarly compulsory, our survey indicates that the vast majority of students attended most of them (4.7 (0.6) CAs, n=132). Interestingly, biomedical background students attended significantly more CAs than students with a non-biomedical background, although attendance was still high for both cohorts (4.8 (0.5) n=93 *vs* 4.6 (0.7), n=42; p=0.02; *data not shown*).

Students' attitude towards summative GM1001 and formative GM1002 CAs

We asked students to respond to seven separate statements that sought to gauge their perception of both summative and formative CAs for their understanding of physiology, with the responses shown in figures 2-4.

Statement 1. **"I used the CAs to provide me with feedback about my understanding of physiological concepts/areas**".

For the comparison of summative *vs* formative CAs, 89.4% (χ^2 = 149, p<0.0001; 4.2 (0.9), n=132) *vs* 90.1% (χ^2 = 106, p<0.0001; 4.4 (0.8), n=131), respectively, of students agreed (A) or strongly agreed (SA) with the statement (Figures 2A&B). Similar overall values in response to the same statement were also obtained when students were subdivided into biomedical *vs* non-biomedical background, *i.e.* summative CAs - biomedical, 85.6% (χ^2 = 58.3, p<0.0001 (4.2 (0.11), n=93; non-biomedical, 90.5% (χ^2 = 17.2, p<0.0001; 4.3 (0.1), n=42): formative CAs – biomedical, 90.2% (χ^2 = 77.2, p<0.0001; 4.4 (0.8), n=94); non-biomedical, 92.7% (χ^2 = 32.4, p<0.0001; 4.3 (0.8), n=41) (*data not shown*). No significant differences were recorded between cohorts.

Statement 2. "**The CAs were a valuable learning exercise**".

For the summative CAs, 74.3% (χ^2 = 88.0, p<0.0001; 4.1 (0.9), n=132) of students agreed or strongly agreed with statement 2, in comparison to 78.6% (χ^2 = 74.2, p<0.0001; 4.1 (1.0), n=131) for the formative CAs (Figures 2A&B).

Similarly, responses of biomedical and non-biomedical background students did not significantly differ from each other with regard to the summative *vs* formative CAs, *i.e.*, summative CAs - biomedical, 73.1% SA or A (χ^2 = 40.0, p<0.0001; 4.0 (1.1), n=93); non-

biomedical, 78.6% SA or A (χ^2 = 11.4, p<0.0001; 4.1 (0.9), n=42): formative CAs – biomedical, 77.4% SA or A (χ^2 = 51.2, p<0.0001; 4.1 (1.0), n=93), non-biomedical, 82.9% SA or A (χ^2 = 26.0, p<0.0001; 4.1 (0.9), n=41)(*data not shown*). No significant differences were recorded between cohorts.

Statement 3. "**The CAs motivated me to study and to keep up with the taught material**", In contrast to the previous two questions, responses to statement 3 differed markedly with respect to sentiment towards the summative *versus* formative CAs. For example, for students overall, for the summative CAs 76.5% A or SA (χ^2 = 88.0, p<0.0001; 4.1 (1.1), n=132), compared to only 59.7% A or SA (χ^2 = 37.7, p<0.0001; 3.7 (1.2), n=130) for the formative CAs (p=0.006, unpaired t-test)(Figures 2A, B & C). However, when we sub-divided responses into those from biomedical *vs* non-biomedical students, the picture was less clear. For example, although both groups were strongly in agreement that the summative CAs motivated them to study and keep up with material, the non-biomedical background students felt this motivation more (*i.e.* non-biomedical - 83.3 % SA or A (χ^2 = 13.0, p<0.0001; 4.3 (0.9), n=42; biomedical - 73.1% (χ^2 = 36.0, p<0.0001; 4.0 (1.2), n=93), although this difference was not statistically significant (p=0.06, unpaired t-test)(Figures 2 A, B & C). However, our results show that both biomedical and non-biomedical background students found the formative CAs less motivating than the summative CAs (*i.e.* biomedical – 64.5 % SA or A (χ^2 = 28.1, p<0.0001; 3.8 (1.1), n=93); non-biomedical 65% SA or A (χ^2 = 21.8, p<0.0001; 3.6 (1.2), n=40). Indeed, for the non-biomedical background students, this decrease in positive sentiment towards statement 3 for the formative, relative to the summative, CAs was statistically significant (non-biomed GM1001 (4.3 (0.9), n=42 *vs* non-biomed GM1002 (3.6 (1.2), n=40; p=0.002, unpaired t-test)(Figures 2C).

Statement 4. "**The CAs helped me to identify deficits in my knowledge of specific areas of physiology"**

Again, overall, there was very positive sentiment towards statement 4 for both summative (77.3% SA or A (χ^2 = 96.4, p<0.0001; 4.0 (1.0), n=132) and formative (83.7% SA or A (χ^2 = 95.9, p<0.0001; 4.2 (0.8), n=131) CAs, the formative CAs were significantly more favored in this regard than the summative CAs (p=0.04, unpaired t-test) (Figures 3A, B&C).

Similarly, the sentiment of biomedical and non-biomedical background students towards the same statement for both summative and formative CAs was equally positive, with both student cohorts also preferring the formative to the summative CAs, significantly so for the biomedical background students, in this particular context (biomedical - summative 73.9% SA or A (χ^2 = 40.0, p<0.0001; 4.0 (1.1), n=93) *vs* formative 83.7% SA or A (χ^2 = 95.9, p<0.0001; 4.3 (0.9), n=93) (p=0.03, unpaired t-test); non-biomedical – summative 79.1% SA or A (χ^2 = 8.0, p<0.0001; 4.2 (0.9), n=42) *vs* formative 87.8% SA or A (χ^2 = 27.9, p<0.0001: 4.1 (0.8), n=41) (*n.s.* unpaired t-test)(Figures 3A, B&C).

Statement 5. "**The results from the CAs helped to direct my study before end of module or end of year exams**"

In contrast to the responses to the preceding statements, there was a significant decrease in positive sentiment from the students overall to statement 5, for both the summative and formative CAs. Furthermore, we also discovered that, in this particular context, the summative CAs were significantly less popular than the formative CAs. Specifically, for the summative CAs, only 55.0% (χ^2 = 29.0, p<0.0001; 3.5 (1.2), n=132) of students, agreed or strongly agreed with statement 5 compared to 64.6% (χ^2 = 54.6, p<0.0001; 3.8 (1.1), n=131) for the formative CAs (p=0.02; unpaired t-test)(Figures 3A, B&D), which mirrors in sentiment the responses to statement 4 about the CAs helping to identify gaps in knowledge.

When we subdivided the students into biomedical or non-biomedical backgrounds, it was notable that the biomedical background students were markedly less enthusiastic than the non-biomedical students about either summative or formative CAs in their ability to direct their study for exams. For example, with regard to the summative CAs alone, 76.3% of non-biomedical students agreed or strongly agreed ($\chi^2 = 9.5$, $p<0.0001$; 3.8 (1.1), n=38) with the statement, whereas only 47.3% of the biomedical students did so ($\chi^2 = 5.8$, $p=0.04$; 3.3 (1.2), n=91) ($p=0.01$, unpaired t-test). Similarly for the formative CAs, 75.6% ($\chi^2 = 23.8$, $p<0.0001$; 4 (0.9), n=41) *versus* 59.8% ($\chi^2 = 20.1$, $p<0.0001$; 3.7 (1.1), n=93) of non-biomedical and biomedical students, respectively, agreed or strongly agreed with the statement, with this difference only just failing to reach statistical significance ($p=0.06$, unpaired t-test) (Figures 3A, B&D).

Statement 6: "**As there weren't any marks awarded for the CAs in GM1002, I didn't feel the need to study for them**".

We found that approximately half of the students, 46.1%, disagreed (D) or strongly disagreed (SD) ($\chi^2 = 10.8$, $p<0.005$; 2.8 (1.3), n=130) with statement 6. It is worth noting however that a sizable minority of the students (30.8%) did also agree or strongly agree with the same statement, with a further 23.1% neither agreeing nor disagreeing. Although more non-biomedical students disagreed or strongly disagreed (53.8% ($\chi^2 = 8.0$, $p=0.02$; 2.7 (1.2), n= 40) with statement 6 than biomedical students (43.5% ($\chi^2 = 4.8$, $p=0.09$; 2.8 (1.3), n=93), the difference was not significant (Figures 4A&B).

Statement 7. "**If summative marks were NOT awarded for the CAs in GM1001, then I would not study for them"**

60.6% of students disagreed or strongly disagreed with statement 7, whereas only a quarter (25.7%) of respondents agreed or strongly agreed with it (χ^2 = 46.9, p<0.0001; 2.5 (1.1), n=132). A slightly greater percentage of biomedical students disagreed or strongly disagreed with the statement than the non-biomedical background students (55.8% (χ^2 = 18.6, p<0.0001; 2.6 (1.1), n=95) *versus* 52.4% (χ^2 = 5.8, p=0.02; 2.5 (0.9), n=42), but this difference was not significant (Figure 4A&B).

Do students prefer formative or summative assessments?

The final part of the survey specifically asked students to indicate what their preferred format of CA was, or would be, from four options (*i.e.* the summative and formative formats that they had experienced in GM1001 and GM1002, respectively, plus two additional possible formative formats); "Formative, completed under test conditions on computers in the physiology lab with answers provided immediately or shortly afterwards", "Formative, completed as an individual or group exercise (*e.g.* using Socrative)", "Formative, with questions placed on [a Learning Management System such as] Blackboard for students to complete in their own time", or "Summative (*i.e.* marks contributing to EOM exam), completed under test conditions on computers in the physiology lab with answers provided immediately or shortly afterwards".

As can be seen in figure 4C, 71.9% (92/128) of students opted for one or other of the three formative CA formats offered, with the formative CA format that they had experienced in GM1002 being the single most popular option (38.3%; 49/128 students). Notably, the summative format experienced in GM1002 was selected by only 28.1% (36/128) of students. It is interesting to note however that although biomedical background students also preferred the GM1002 formative CA format to the GM1001 summative CA format (41.4% (36/87

students) *vs* 21.8% (19/87 students), with 78.2% (68/87) opting for formative CAs generally, this trend was reversed with the non-biomedical background students. Specifically, a much smaller percentage of students from non-biomedical backgrounds than students overall opted for formative assessments generally (58.6%; 24/41 students), with a greater proportion of those same students preferring the summative CAs used in the study (41.5%; 17/41 students) to those that preferred the GM1002-style formative CAs (29.3%; 12/41 students; p=0.003 *vs* biomedical background students)(Figures 4C & D).

Discussion

In line with the relatively few studies investigating pre-clinical medical students' assessment-type preferences, the current study revealed that the sentiment of UCC pre-clinical medical students towards regular testing of their knowledge of physiology, using either frequent summative or formative formats, was overwhelmingly positive in terms of being both valuable learning exercises and providing feedback about their understanding of physiology (statements 1 and 2). Additionally, and in line with the assertion that "assessment [can] drive learning" (40, 41) most students indicated that both formats of CA motivated them to study and learn the taught material (statement 3). However, our results show that the summative tests were significantly more effective in this regard than the no-stakes, formative CAs. Interestingly, the greatest difference in the degree of motivation provided by summative *versus* formative CAs was with the non-biomedical background students who were both significantly more motivated than biomedical background students to study for summative CAs, but significantly *less* motivated to study for the formative CAs. However, more students from the non-biomedical than the biomedical cohort, disagreed with statement 6 that, because marks weren't available for the formative tests, then they didn't study for them.

- *"the formative nature of* [the GM1002 CAs] *slightly removed my motivation to study for them, in contrast to* [the GM1001 CAs]*, where I was more motivated to revise for the CAs, in spite of their small contribution to the overall mark"*

- *"I think that the summative CAs - regardless of how negligible their contributions to the final grade - will always incentivise students to acquire marks instead of the knowledge"*

Similarly, although overall more than 60% of students indicated that they would study for the summative tests even in the absence of marks being available, when asked about their actual study habits for the formative tests, over half of the class were, at best, non-committal on the matter.

- "*I felt more pressure to learn and know the material in GM1001 but I preferred the less stressful aspect of GM1002. Probably overall having* [summative CAs] *contribute to marks is better so that we are encouraged to keep up with the large amount of material.*"

However, in spite of the reduced motivation to study for the formative *versus* summative CAs, the formative testing was more popular with the students for identifying deficits in their knowledge of physiological topics than the summative CAs (statement 4). Although we did not formally interrogate this finding with students, anecdotally, the immediate feedback provided on each question during the formative CAs, whereby any misapprehensions or misunderstandings can be clarified in real time, was one of the main reasons for students' preference.

- "*I preferred having solutions provided immediately (eg via Socrative), because it either reinforced what I knew OR I could correct myself on the spot (compared to reviewing CAs a week or two later, when I can't remember how I answered it* [sic] *and have moved on from reviewing that material in my own studies*"

This preference for formative over summative testing overall was confirmed when they were asked directly to select their favored test format, with 71.8% of students preferring some form

of formative testing over summative tests, with the largest proportion, 38.3% (49/128), opting for the format of formative CAs utilized in the current study.

- *"I like this option because you address the answer to the question right away and there is opportunity for discussion with lecturer/peers even if you prefer to answer alone. It also makes for better revision"*

- *"I learned more from doing the tests on socrative, where I could save the answers and explanations to review later. The lab-based assessment was not useful, particularly at the beginning of the year, when everything was so new - it added undue stress*

- "*I used the CAs to motivate me to stay on top of my work and see how well I knew the content at that point and time. I would not want the CAs to be worth more than 1% because I wouldn't want the stress of prioritizing the CAs over other subjects. Therefore I think the group Socrative* [sic] *had the best outcome of pushing me to stay on top of my work without the stress of a summative CA*"

However, one interesting feature that arose from these particular data was the finding that, in contrast to all of the other cohorts analysed, non-biomedical background students not only had a much lower preference for formative assessments in general (58.6%), but also contained a greater proportion of students (compared to biomedical background, and students overall) that preferred the style of summative CAs used in GM1001 (41.5%) to those that preferred the GM1002-style formative CAs (29.3%). Notably, non-biomedical background students also attended significantly fewer of the formative CAs than those students with a biomedical background (*data not shown*). The reason for such a marked preference for summative CAs in

this particular cohort is unclear. One possibility is that they disliked the collaborative-style formative CA format to that of the summative CAs, which were sat individually, under proctored conditions (non-biomedical background students indicated a significant preference for working alone rather than in groups (for the formative CAs) compared to the biomedical background students (*data not shown*)), perhaps because it might have revealed a weaker physiology knowledge to their peers. However, this explanation alone seems unlikely as only 7.3% of non-biomedical students selected the "Formative, completed under test conditions on computers in the physiology lab, with answers provided immediately or shortly afterwards" option for that particular survey item. Another possibility is provided by responses to an earlier survey item, whereby non-biomedical background students were more positive towards the summative CAs than their biomedical peers in motivating them "to study and keep up with taught material". Similarly, a significantly larger proportion of non-biomedical than biomedical background students indicated that the summative CAs (76.3% *vs* 47.3%, respectively) helped to direct their study prior to formal exams, although the relative percentages for non-biomedical *versus* biomedical background students were very similar with regard to the formative CAs for the same survey item.

Limitations of study

One limitation of this study is that all of the data were obtained from a specific cohort of students attending one particular medical school, limiting the potential to extrapolate the findings to other medical programs. However, by conducting the study across the two years, and ruling out as many confounding factors as possible, such as lecturer to lecturer, school to school, and program to program variability, as well as keeping the delivery of material and the two modes of CAs identical for the whole study, our study ensured that students' responses

were specifically focussed on their perceptions of the way that the summative and formative CAs had been delivered, rather than being potentially influenced by these other factors.

A further criticism of the study could be that as the summative assessments were delivered in a proctored format, with each student answering questions individually, whereas it was possible for students to discuss and answer questions as a group for the formative CAs, thereby making direct comparisons between the two modes of CA delivery invalid. However, whilst this criticism is valid to a certain extent, we feel that fact that only 32.6% of respondents indicated that they had answered the formative CAs as part of group of two or more, *i.e.* more than two thirds of the respondents had completed the quizzes individually as they had done for the summative CAs, means that reasonable comparisons can be drawn between the data for the two CA types.

A further limitation is that there may have been survey bias towards the formative CAs due to the fact that, for both years of the survey, it was the most recently experienced of the two CA formats. To counter such possible bias, it would have been useful to have flipped the summative and formative semesters in year two of the survey.

Although we did provide a free-response comment box at the end of the survey for students to enter any feedback on the survey questions that they felt was relevant (with some of the quotes provided being included above), an essay-type survey question where students were specifically asked to address why they preferred one mode of CA over another, such as the lack of peer assistance, delayed feedback or the more formal atmosphere of the summative CAs, would have also been useful. Furthermore, it would also have been interesting to specifically determine why non-biomedical students preferred summative CAs over formative CAs, in direct contrast to biomedical students and the entire student cohort as a whole. Unfortunately,

as we did not anticipate such distinct differences in opinion between these cohorts, it was not factored into the original survey design.

Finally, due to the anonymous nature of the survey and potential ethical issues associated with this factor, we were unable to set up the study in such a way as to determine if either the summative or formative CAs had any significant effect on exam performance. However, previous studies have shown that regular formative testing significantly improved grades in a 1st year biology course (4).

Conclusions

Although our results indicate that pre-clinical medical students as a whole generally prefer formative to summative in-term assessments (notwithstanding the preferences of students from non-biomedical backgrounds), it is clear that they are aware of the relative benefits of both formats. Specifically, formative assessments provide immediate feedback on taught material in real time, in a relatively informal and relaxed environment, whereas summative assessments not only increase the motivation of students to keep up with coursework, but also test their knowledge in a lower-stakes examination environment that helps to prepare them for future, more formal, higher stakes assessments. As such, and in line with the previous proposal of Dixson and Worrell (2016), as well as reaping the benefits of the testing effect (12, 42), students are likely to benefit from a healthy balance of regular formative and summative assessments due to the fact that both forms of assessment should complement each other: the former facilitating understanding, and the latter evaluating the degree of understanding at the end of a defined unit of taught material. This is supported by other studies which have found that a combination of formative and summative assessments can be more effective than either type of assessment alone (43, 44).

References

1. **Slominski T, Grindberg S, and Momsen J**. Physiology is hard: a replication study of students' perceived learning difficulties. *Advances in Physiology Education.* 43: 121-127, 2019. https://doi.org/10.1152/advan.00040.2018
2. **Dixson DD, and Worrell FC**. Formative and summative assessment in the classroom. *Theory Into Practice* 55: 153-159, 2016. https://doi.org/10.1080/00405841.2016.1148989
3. **Garrison C, and Ehringhaus M**. Formative and summative assessments in the classroom (Online). Westerville, OH: AMLE. https://www.amle.org/BrowsebyTopic/WhatsNew/WNDet/TabId/270/ArtMID/888/ArticleID/286/Formative-and-Summative-Assessments-in-the-Clasrrom.aspx [21 Dec 2022].
4. **Bailey E, Jensen J, Nelson J, Wiberg H, and Bell J**. Weekly formative exams and creative grading enhance student learning in an introductory biology course. *CBE—Life Sciences Education* 16: ar2, 2017. https://doi.org/10.1187/cbe.16-02-0104
5. **Bangert-Drowns RL, Kulik C-LC, Kulik JA, and Morgan M**. The instructional effect of feedback in test-like events. *Review of Educational Research* 61: 213-238, 1991. https://doi.org/10.3102/003465430610022
6. **Black P, and Wiliam D**. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice* 5: 7-74, 1998. https://doi.org/10.1080/0969595980050102
7. **Kibble J**. Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiology Education* 31: 253-260, 2007. https://doi.org/10.1152/advan.00027.2007
8. **Kitchen E, Bell JD, Reeve S, Sudweeks RR, and Bradshaw WS**. Teaching cell biology in the large-enrollment classroom: methods to promote analytical thinking and assessment of their effectiveness. *Cell Biology Education* 2: 180-194, 2003. https://doi.org/10.1187/cbe.02-11-0055
9. **Panadero E, and Jonsson A**. The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review* 9: 129-144, 2013. https://doi.org/10.1016/j.edurev.2013.01.002
10. **Preston R, Gratani M, Owens K, Roche P, Zimanyi M, and Malau-Aduli B**. Exploring the Impact of Assessment on Medical Students' Learning. *Assessment & Evaluation in Higher Education* 45: 109-124, 2020. https://doi.org/10.1080/02602938.2019.1614145
11. **Rushton A**. Formative assessment: a key to deep learning? *Medical Teacher* 27: 509-513, 2005. https://doi.org/10.1080/01421590500129159
12. **Roediger HL, 3rd, and Karpicke JD**. The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science* 1: 29181, 2006. https://doi.org/10.1111/j.1745-6916.2006.00012.x
13. **Anwar M, and Hameed FM**. MEDICAL EDUCATION SECTION: Students' Perceptions towards Formative and Summative Assessment: A Single Institution Study. *Journal of Islamic International Medical College (JIIMC)* 11: 35-40, 2016. https://doi.org/10.1111/j.1745-6916.2006.00012.x
14. **Brookhart SM**. A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education* 10: 161-180, 1997. https://doi.org/10.1207/s15324818ame1002_4
15. **Brookhart SM**. Expanding views about formative classroom assessment: A review of the literature. *Formative Classroom Assessment: Theory into Practice* 43-62, 2007. https://doi.org/10.1080/0969594X.2010.513678
16. **McMillan JH**. *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson, 2014.

17. **Birenbaum M, and Feldman RA**. Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research* 40: 90-98, 1998. https://doi.org/10.1080/0013188980400109
18. **Scouller K**. Influence of assessment method on students' learning approaches, perceptions and preferences: The assignment essay versus the short answer examination. *HERDSA 1996 conference proceedings*. 1996. Retrieved 21 Dec 2022, from http://www.herdsa.org.au/confs/1996/scouller.html.
19. **Scouller K**. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education* 35: 453-472, 1998. http://www.jstor.org/stable/3448270
20. **Segers M, Nijhuis J, and Gijselaers W**. Redesigning a learning and assessment environment: The influence on students' perceptions of assessment demands and their learning strategies. *Studies in Educational Evaluation* 32: 223-242, 2006. https://doi.org/10.1016/j.stueduc.2006.08.004
21. **Van de Watering G, Gijbels D, Dochy F, and Van der Rijt J**. Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education* 56: 645-658, 2008. https://www.jstor.org/stable/40269095
22. **Cheng L, Wu Y, and Liu X**. Chinese university students' perceptions of assessment tasks and classroom assessment environment. *Language Testing in Asia* 5: 13, 2015. doi:10.1186/s40468-015-0020-6
23. **Norcini JJ, and McKinley DW**. Assessment methods in medical education. *Teaching and Teacher Education* 23: 239-250, 2007. https://doi.org/10.1016/j.tate.2006.12.021
24. **Al-Wardy NM**. Assessment methods in undergraduate medical education. *Sultan Qaboos University Medical Journal* 10: 7, 2010. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074721/pdf/squmj-10-2-203.pdf
25. **Amin TT, Kaliyadan F, and Al-Muhaidib NS**. Medical students' assessment preferences at King Faisal University, Saudi Arabia. *Advances in Medical Education and Practice* 2: 95-103, 2011. https://doi.org/10.2147/AMEP.S12950
26. **Brookhart SM**. Classroom Assessment: Tensions and Intersections in Theory and Practice. *Teachers College Record* 106: 429-458, 2004. https://doi.org/10.1111/j.1467-9620.2004.00346.x
27. **Abdulla MH, Brint E, and Rae MK**. Teaching physiology to medical students in the COVID-19 era with synchronous formative assessments utilizing simultaneous, combined Zoom and Socrative platforms. *Scierea Journal of Education* 6: 12-32, 2021. https://doi.org/10.54647/education88192
28. **Rae MG, and O'Malley D**. Using an online student response system, Socrative, to facilitate active learning of Physiology by first year graduate entry to medicine students: a feasibility study. *MedEdPublish* 6: 2017. https://mededpublish.org/articles/6-4
29. **Beller M, and Gafni N**. Can Item Format (Multiple Choice vs. Open-Ended) Account for Gender Differences in Mathematics Achievement? *Sex Roles* 42: 1-21, 2000. https://doi.org/10.1023/A:1007051109754
30. **Ben-Chaim D, and Zoller U**. Examination-type preferences of secondary school students and their teachers in the science disciplines. *Instructional Science* 25: 347-367, 1997. http://www.jstor.org/stable/23371248
31. **Krathwohl DR**. A revision of Bloom's taxonomy: An overview. *Theory Into Practice* 41: 212-218, 2002. https://doi.org/10.1207/s15430421tip4104_2
32. **Awedh M, Mueen A, Zafar B, and Manzoor U**. Using Socrative and Smartphones for the support of collaborative learning. *arXiv preprint arXiv:150101276* 2015. https://doi.org/10.48550/arXiv.1501.01276
33. **Coca DM, and Ciesielkiewicz M**. The Use of Smartphones in Primary Education: Improving Arithmetic Calculations in Pre-service Primary Teacher Training. *US-China Education Review* 403, 2014.

34. **Dervan P**. Increasing in-class student engagement using Socrative (an online Student Response System). *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education* 6: 1801-1813, 2014. https://ojs.aishe.org/index.php/aishe-j/article/view/180/283
35. **Kaya A, and Balta N**. Taking Advantages of Technologies: Using the Socrative in English Language Teaching Classes. *International Journal of Social Sciences & Educational Studies* 4, 2016.
36. **Liu DYT, and Taylor CE**. Engaging students in large lectures of introductory biology and molecular biology service courses using student response systems. In: *Proceedings of The Australian Conference on Science and Mathematics Education (formerly UniServe Science Conference)* 2013.
37. **Wash PD**. Taking advantage of mobile devices: Using Socrative in the classroom. *Journal of Teaching and Learning with Technology* 3: 99-101, 2014. https://doi.org/10.14434/jotlt.v3n1.5016
38. **McMaster KL, Fuchs D, and Fuchs LS**. Research on peer-assisted learning strategies: The promise and limitations of peer-mediated instruction. *Reading & Writing Quarterly* 22: 5-25, 2006. https://doi.org/10.1080/10573560500203491
39. **Menezes A, Burgess A, Clarke AJ, and Mellis C**. Peer-assisted learning in medical school: tutees' perspective. *Advances in Medical Education and Practice* 7: 31, 2016. http://dx.doi.org/10.2147/AMEP.S94570
40. **McLachlan JC**. The relationship between assessment and learning. *Medical Education* 40, 716–777, 2006. https://psycnet.apa.org/doi/10.1111/j.1365-2929.2006.02518.x
41. **Muijtjens AMM, Hoogenboom RJI, Verwijnen GM, van der Vleuten CPM**. Relative or Absolute Standards in Assessing Medical Knowledge Using Progress Tests. *Advances in Health Sciences Education* 3: 1573-1677. https://doi.org/10.1023/A:1009728423412
42. **Larsen DP, Butler Ac Fau - Roediger HL, 3rd, and Roediger HL, 3rd**. Test-enhanced learning in medical education. *Medical Education* 42: 7, 2008. https://doi.org/10.1111/j.1365-2923.2008.03124.x
43. **Black P, and Wiliam D**. Inside the black box: Raising standards through classroom assessment. *Phi delta kappan* 92: 81-90, 2010. https://doi.org/10.1177/003172171009200119
44. **Sadler DR**. Formative assessment and the design of instructional systems. *Instructional Science* 18: 119-144, 1989. https://doi.org/10.1007/BF00117714

Figure Legends

Figure 1. Student demographics. Pie charts illustrating distribution of 1st year graduate entry to medicine (GEM) survey respondents in relation to sex (A), age (B), primary degree background (C) and region of origin (D) (EU = European Union) for academic years 18-19 and 19-20 (A).

Figure 2. Students' views on statements #1-3. Students provided their views on three statements using a five point Likert scale ranging from 5 (strongly agree) to 1 (strongly disagree). Graph shows Likert scale data presented as mean scores for overall student responses to statements 1-3 for summative and formative, with error bars representing standard deviation. *Inset:* The percentage of positive responses, "agree" or "strongly agree" for each statement for both summative and formative CAs (A). Graph showing Likert scale data for summative and formative CAs, specifically in relation to statement # 3. Responses are shown for students overall, students with a biomedical primary degree and those with a non-biomedical primary degree. Likert scale data presented as mean scores for overall student responses to statement # 3 with error bars representing standard deviation (B). ** indicates $p<0.01$.

Figure 3. Students' views on statements #4 & #5. Students provided their views on statements, #4 and #5, using a five point Likert scale ranging from 5 (strongly agree) to 1 (strongly disagree). Graph shows Likert scale data presented as mean scores for overall student responses to the statements for summative (filled circles) and formative (filled squares) CAs, with error bars representing standard deviation. Inset: The percentage of positive responses, "agree" or "strongly agree" for each statement for both summative and formative CAs (A). Graph showing Likert scale data for summative and formative CAs, specifically in relation to statement #4. Responses are shown for students overall, students with a biomedical primary degree and those with a non-biomedical primary degree (B). Graph showing Likert scale data for summative and formative CAs, specifically in relation to statement #5. Responses are shown for students overall, students with a biomedical primary degree, and those with a non-biomedical primary degree (C). * indicates $p<0.05$, ** indicates $p<0.01$.

Figure 4. Students' views on statements #6 & #7 and preference for summative or formative CAs. Students provided their views on a further two statements, #6 and #7, using a five point Likert scale ranging from 5 (strongly agree) to 1 (strongly disagree). Graph shows Likert scale data presented as mean scores for overall student responses to the statements for summative (filled circles) and formative (filled squares) CAs, with error bars representing SD. *Inset:* The percentage of positive responses, "agree" or "strongly agree" for each statement (A). Graph illustrating the percentage student preferences for GM1001 summative CAs *versus* the format of formative CAs used in GM1002. Responses are

shown for students overall, students with a biomedical primary degree and those with a non-biomedical primary degree. ** indicates $p<0.01$ (B). Pie chart illustrating students' percentage preference for summative or formative CA formats (C).