Received: 13 September 2022 | Revised: 23 June 2023 | Accepted: 24 June 2023

DOI: 10.1002/ca.24092

EDUCATION

CLINICAL ANATOMY WILEY

Spot(ters) the difference: Bringing traditional anatomical examinations online

Alistair P. Bond[1] | Thomas Butts[2] | Claire M. Tierney[1]

[1]Human Anatomy Resource Centre, Education Directorate, Faculty of Health and Life Science, University of Liverpool, Liverpool, UK

[2]School of Medicine, University of Sunderland, Sunderland, UK

Correspondence
Alistair P. Bond, Human Anatomy Resource Centre, University of Liverpool, Liverpool, UK.
Email: apbond@liverpool.ac.uk

Abstract

The COVID-19 pandemic caused a shift in anatomy education forcing institutions to find innovative ways to teach and assess online. This study details the development of an online spotter across multiple modules that allowed students to sit the examination at home whilst still maintaining the integrity of the assessment. The online spotter consisted of individual, Zoom calls between students and examiners whereby slides with images and questions were screen shared. To examine the viability of this spotter in non-lockdown scenarios several parameters were considered. Mean marks were compared to traditional versions and Pearson's *r* correlation coefficients were calculated between online and traditional spotters and between online spotters and overall performance in anatomy modules. A survey was carried out to determine the students' view of the assessment. Pearson's *r* was between 0.33 and 0.49 when comparing online spotters to the traditional format, and between 0.65 and 0.75 ($p < 0.01$) when compared to a calculated anatomy score. The survey indicated overall student satisfaction as 82.5% reported that it was a fair way to test their knowledge and 55% reported the same or lower levels of anxiety when compared to traditional spotters. However, there was nothing to indicate that the students preferred this format over laboratory-based spotters. These results indicate that this new exam format would be useful for small cohorts who are undertaking online or hybrid courses, or in circumstances when running a full spotter is too costly, and represents a fair and robust way to assess practical anatomical knowledge online.

KEYWORDS
anatomy assessment, COVID-19, online anatomy, remote assessment, spotter, steeplechase

1 | INTRODUCTION

The COVID-19 pandemic forced a seismic change in anatomy education as many institutions were forced to deliver teaching in a virtual and online environment (Brassett et al., 2020). Anatomy departments utilized a variety of methods to ensure that learning objectives were delivered to students; these include using 3D anatomical software, photogrammetry generated prosections, online quiz packages, and teaching in virtual classrooms (Attardi et al., 2021).

Examinations were also forced to adapt during the pandemic lockdown, with many altered to allow students to sit them from home, or in some cases canceled altogether (Harmon et al., 2021). Assessments in anatomy are key to the learning process and have been suggested to be more important to student learning than the format and style of how teaching is delivered (Alraddadi et al., 2021). Spotter, or

steeplechase, exams are commonplace in anatomy and allow for the examination of specific learning objectives that are difficult to test in written examinations. For example, learning objectives that use verbs such as find, identify, locate, recognize, and label can all be assessed more effectively in a spotter. Despite the circumstances of the lockdown, and to comply with the module specifications, these objectives still had to be examined during the pandemic.

The spotter examination is a station-based assessment where students are presented with a specific question related to a prosection, model or medical image on each station. Each station usually lasts 1 min and has two parts; a basic identify question and a functional/developmental/clinical/evolutionary question to test the higher levels of Bloom's taxonomy (Alraddadi et al., 2021; Merzougui et al., 2021; Smith & McManus, 2015).

The spotter is regarded as a gold standard in assessing practical anatomy and is the most recommended form of assessment by medical students, trainees and specialists (Rowland et al., 2011). Spotters test a student's 3D spatial understanding (Brenner et al., 2015; Smith & McManus, 2015) and are better than using multiple choice questions (MCQs) alone (Schubert et al., 2009). They help to test students' knowledge of structural relationships, ability to differentiate between two similar structures (Alraddadi et al., 2021) and recognition of anatomical variants (Brenner et al., 2015; Smith & McManus, 2015).

Spotters are not only championed by teachers; students also acknowledge that, while stressful, they are the most effective way to test anatomical knowledge (Merzougui et al., 2021). Despite falling numbers of spotters in anatomy education, likely due to practical considerations rather than educational matters, both undergraduate and postgraduate medical students have stated in a study by Rowland et al. (2011) that they prefer practical based exams.

As with any assessment, spotters do have drawbacks, for example they can be costly to run, in specimen preparation, staff time and use of facilities (Choudhury et al., 2016). They are also argued to only test the lower levels of Bloom's taxonomy (Smith & McManus, 2015) and, although using pen and paper in spotters is decreasing, it can be difficult to mark handwritten answers (Choudhury et al., 2016). These must all be considered when setting a spotter examination as part of an overall assessment strategy.

Given the importance of the spotter examinations, and the need to properly assess learning objectives covered by these practical based exams during the pandemic, an online spotter was devised using Zoom video conferencing software (Zoom Video Communications, 2022). These online spotters were implemented in four modules—two in first year and two in second year—of the Anatomy and Human Biology degree at the University of Liverpool.

By considering the implementation of these online spotters, this paper aims to discuss the viability of the exam format in relation to other, non-lockdown, circumstances. This is important to consider given that there is an ever-increasing online approach to teaching (Guimaraes et al., 2018) which raises the possibility of more anatomy courses being taught online or in a hybrid manner. Indeed, it was noted as long as 10 years ago that online teaching was expanding but assessment was not keeping up (Inuwa et al., 2012). The online spotter presented here may be one way of ensuring the spotter can stand the test of time and keep its position as the gold standard of practical anatomy assessment.

2 | MATERIALS AND METHODS

2.1 | Cohorts and timeline

During the COVID-19 UK lockdown in March 2020 online spotter examinations were implemented for the first-year module 'Circulatory and Respiratory Anatomy'(LIFE116) and two second-year modules 'Anatomy of the Head and Neck'(LIFE220) and 'Functional Neuroanatomy'(LIFE218). The academic year 2020–2021 saw the addition of online spotters for the 'Human Locomotor System'(LIFE219) and 'Anatomy of the Abdomen and Pelvis'(LIFE235). In total, over a year and a half, eight online spotters were delivered (Table 1 and Figure 1).

2.2 | Online spotters

Each online spotter consisted of 16–24 PowerPoint slides containing an image of a prosection, a diagram, or a medical image. These images were sourced from either 'Gray's Anatomy for Students' (Drake et al., 2020), 'Clinical Atlas of Human Anatomy' (Abrahams et al., 2020), or an internal library. An arrow would indicate a specific structure and the slide contained a question.

To begin the exam, the student would log on to Zoom at a predetermined time with their camera on. A member of academic staff would be hosting the call as the examiner and would be able to

TABLE 1 Module codes, names, and number of students taking the online spotter exams for each module during the COVID-19 pandemic.

		Student numbers	
Module codes	Module names	2019–2020	2020–2021
LIFE 116	Circulatory and Respiratory Anatomy	37	32
LIFE 218	Functional Neuroanatomy	33	33
LIFE 219	Human Locomotor System	n/a	33
LIFE 220	Anatomy of the Head and Neck	33	34
LIFE 235	Anatomy of the Abdomen and Pelvis	n/a	33

Note: Modules with n/a were taken in first semester and were traditional spotter exams.

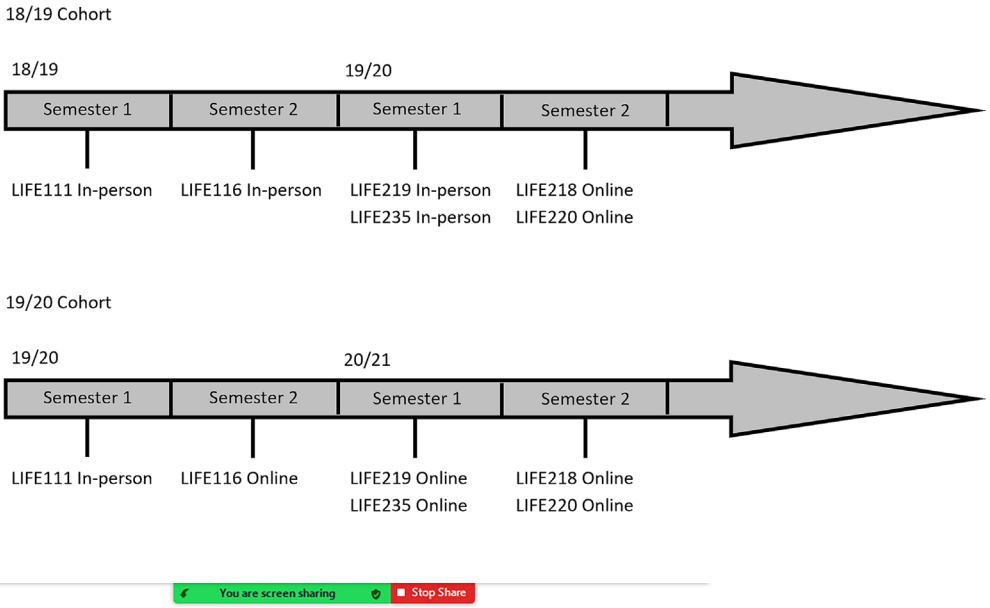

FIGURE 1 A timeline of the spotters taken by each cohort during their undergraduate 'Anatomy and Human Biology' degree at the University of Liverpool. Spotter type; in-person or online, is indicated next to the module code for each cohorts' first 2 years on the degree course.

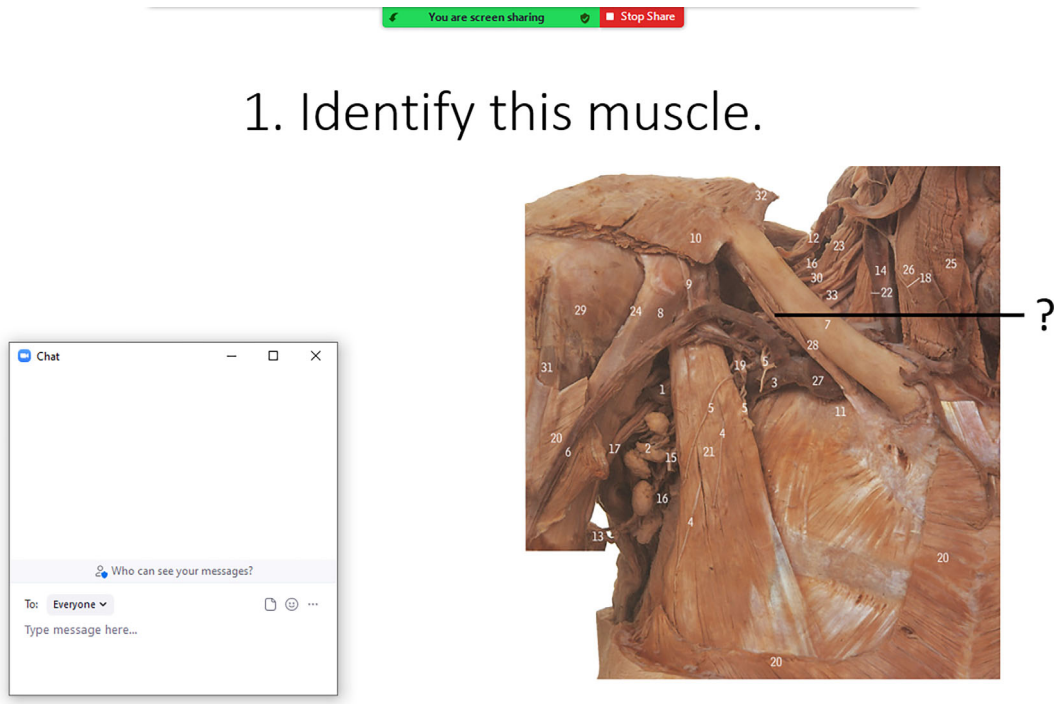

FIGURE 2 A screen shot of a 'Zoom' call with an example question. The question and question number appear at the top of the slide. The image with indicating arrow appear on the right of the screen. The chat box is situated on the left. Image taken from Abrahams et al. (2020).

confirm the identity of the student and ensure that exam conditions were met, such as ensuring they were alone in the room. The examiner would share their screen containing the exam instructions on a PowerPoint slide, which would also be read out. Students were advised to open their chat box and place it in the bottom left of the screen as the slides had been designed in such a way that this area was blank (see Figure 2). Students were offered three options: for the examiner to read out the entire question, the question number alone, or to stay silent throughout the exam. After a final check, with the student's permission, the examiner starts to record the call and begins the exam.

During the exam, students could opt to say their answer out loud to the examiner, type their answer in the chat box function, or do both. Each question was on screen for 30 s in first year modules and 45 s for second year modules. Differences here were due to the individual lecturers setting the exams; 30 s per part is in line with traditional spotters (1 min per station), however, some increased to 45 s to account for the unknown of running the online version. Questions were written to stay in line with traditional, in-person spotters with each question containing a part A for basic identify questions and a part B for advanced, higher order thinking. Different structures were indicated for part B to prevent students being disadvantaged twice if

they did not recognize the structure in the part A. The allotted time for each slide was altered for students who normally received extra time in their exams as a result of pre-existing student support assessments.

Nine members of academic staff were available to examine a maximum class size of 40. Four question sets were written to prevent sharing of questions after the early students had sat the exam. These question sets were written and moderated to ensure fairness across all the student cohort. A question set was also written and on standby in case any students had hardware or internet connection failures so they could be examined at a later date.

On completion of each assessment the examiner sent the video and transcript to the module lead. The transcript was graded against a pre-designed, moderator checked, answer sheet to limit ambiguity and the video was only reviewed in cases of uncertainty.

2.3 | Student preparation

Prior to sitting the exam, students were given instructions on the exam format and the practicalities of using Zoom to sit the exam. Students were given group practice sessions to familiarize themselves with the new format. One week before the exam was due to take place, each student received an Outlook calendar invite with a unique Zoom link from their examiner.

2.4 | Questionnaire

To gain feedback for improvement and get a sense of student acceptability of the exam format, an 11-question survey was sent out to students after they had sat and received the results of their LIFE116, LIFE218 and LIFE220 exams. They had 1 month to complete this. The questions asked can be found in Table 2 with response type listed. Discrete scales were different depending on the question asked.

2.5 | Ethics

Ethical approval was gained from the University of Liverpool Health and Life Sciences Research Ethics Committee. Application number 7876.

2.6 | Statistical comparison of online and in-person spotters

Average spotter scores and standard deviations were calculated for each module for the last 5 years. A two-way ANOVA test was performed to determine statistical differences between the cohorts or spotter format.

TABLE 2 Question number, question asked and percentage responses obtained for student survey.

	Question	Response %				
1	Which year group are you in?	First	Second			
		43	57			
2	Do you believe the Online Spotter Examination was a fair way to test your knowledge?	Yes	No			
		83	17			
3	How good were the instructions provided before the examination in informing you about the format of the spotter?	Very Poor	Poor	Good	Very Good	
		0	8	20	72	
4	How good were the practice sessions provided before the examination in preparing you for the spotter	Very Poor	Poor	Good	Very Good	
		2	15	48	35	
5	Were your anxiety levels higher or lower than that of previous spotter examinations?	Much Lower	Lower	Same	Higher	Much Higher
		2	33	20	22	23
6	Did you feel the length of time received for each question was fair?	Yes	No			
		73	27			
7	Did you prefer to type your answers, give them orally, or both?	Typed	Orally	Both		
		57	5	38		
8	Were there any aspects of the exam which you felt were unfair?	Free response				
9	What could be done to improve online spotter examinations?	Free response				
10	How would you compare the online spotters to the normal spotters that you have experienced?	Free response				
11	Please add any the comments regarding the online spotter examination that you feel would be useful.	Free response				

To determine reliability, online spotters were compared to the traditional spotters for each cohort. A Pearson's correlation coefficient was calculated for each online test; for the 19–20 cohort the online assessments were compared against LIFE111, a lab-based spotter taken in year 1 semester 1. For the 18–19 cohort a mean 'traditional spotter mark' was created for each student from four previous traditional spotters and compared against LIFE218 and LIFE220.

Further Pearson's correlation coefficients were calculated between the online spotter examination and a student's overall performance in anatomy. For all students, an average 'anatomy mark' was calculated from all anatomy modules taken (six in total) and compared against the online spotter scores. *r* Values alongside sample sizes were used to determine significance levels for each comparison.

3 | RESULTS

3.1 | Comparisons of traditional and online formats

The mean spotter scores are shown in Figure 3 for the last five academic years. Those bars that are striped represent spotters that were carried out online due to the COVID-19 pandemic lockdowns. Four out of the eight online spotters achieved the highest scores in the previous 5 years of that module, however there was no significant difference noted between any years for any module.

The 19/20 cohort, when compared against their only traditional spotter had a Pearson's *r* ranging between 0.33 (LIFE116) and 0.42 (LIFE220) (Table 3). The 18/19 cohort had four previous traditional spotters from which to draw upon. Both LIFE218 and LIFE220 achieved a score of 0.49 (Table 4). The online scores were also correlated against a student's overall anatomy score. The Pearson's *r* here ranged from 0.65 in LIFE116 (19/20 cohort) to 0.75 in LIFE218 (18/19 cohort) and LIFE235 (19/20 cohort). All comparisons to the overall anatomy score for both cohorts were significant ($p < 0.01$) (Tables 3 and 4).

3.2 | Questionnaire

Forty students out of a possible 70 completed the questionnaire (response rate = 57%), which comprised of 17 first year students and 23 second year students.

Eighty three percent of students surveyed believed that the online spotter assessment was a fair way to test their anatomical

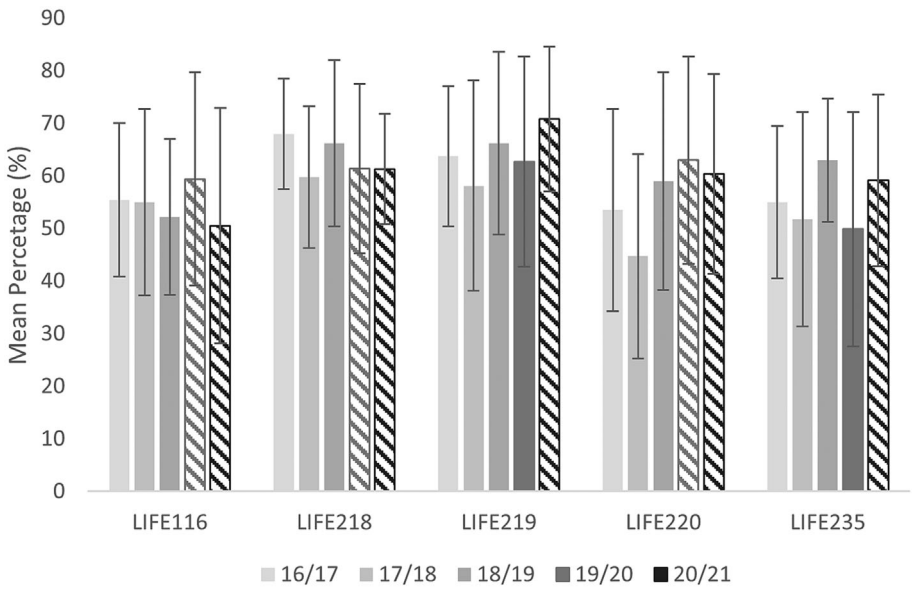

FIGURE 3 Mean spotter results for anatomy modules from academic year 16/17 to 20/21. Striped bars represent those spotters taken online. Error bars represent standard deviation.

TABLE 3 The mean spotter mark, the Pearson's *r* compared to LIFE111 (a traditional spotter) and the Pearson's *r* against an average anatomy score for each online spotter taken by the 19/20 cohort.

	Mean mark (%)	Pearson's *r* versus traditional spotter	Pearson's *r* versus average anatomy score
LIFE116	59	0.33*	0.65**
LIFE218	61	0.4*	0.7**
LIFE219	71	0.31	0.69**
LIFE220	61	0.42*	0.68**
LIFE235	59	0.35*	0.75**

Note: *$p < 0.05$; **$p < 0.01$.

TABLE 4 The mean spotter mark, the Pearson's *r* compared to an average of traditional spotters and the Pearson's *r* against an average anatomy score for each online spotter taken by the 18/19 cohort.

	Mean mark (%)	Pearson's *r* versus spotter score	Pearson's *r* versus average anatomy score
LIFE218	62	0.49**	0.75**
LIFE220	63	0.49**	0.66**

Note: *$p < 0.05$; **$p < 0.01$.

knowledge. More than half (55%) had the same or lower anxiety levels when compared to traditional spotters. Twenty three percent of students had 'much higher' anxiety levels when sitting the online spotter. Seventy three percent of students thought that the time given for each question was fair. However, this drops to 58.5% when looking at first year students (who had 30 s per slide).

Ninety three percent of students felt that the instructions provided before the exam took place were either good or very good, while 82.5% of students felt that the practice sessions before the exam were worthwhile.

Free responses for the questions listed as 'were there any aspects of the exam that you felt were unfair?' and 'what can be done to improve the spotter examinations?' were grouped together as responses for both overlapped. Care was taken not to list the response twice if the same student had made similar comments for both questions. The responses below are listed by the frequency of replies.

Ten responses discussed the need for more questions to prevent each question being too heavily weighted towards the final score. Eight responses reported that there were no aspects that they felt were unfair or no improvements could be made. Six students felt the images were either hard to visualize, or not clear. Four students commented on the short amount of time afforded to each question and three students noted that they felt under a large amount of pressure with the examiner watching. Some improvements were suggested in these comments: one student felt it would help to have the ability to zoom in on the picture, one felt videos would be fairer, and one thought a timer on the screen would have helped them.

For the question 'How would you compare the online spotters to the normal spotters that you have experienced?' the following responses were received.

Eight students reported that anxiety levels were lower for this style of exam, with two making particular mention of being more 'comfortable' at home. Three students noted that they preferred the normal spotter, with three responses mentioning higher anxiety levels and three responses also mentioned that they felt that looking at pictures was not testing the same skills as a normal spotter exam. Three students said the online test was easier while one responder said it was harder.

4 | DISCUSSION

Whilst 2D spotters have been implemented previously (Alraddadi et al., 2021; Brenner et al., 2015; Daly, 2010; Inuwa, Al Rawahy, et al., 2011; Meyer et al., 2016), this specific form of anatomy assessment, using online platforms, has never, to our knowledge, been performed before as a summative test. During the COVID-19 pandemic it allowed for the assessment of learning objectives, such as 'find', 'locate', 'demonstrate' or 'relate', usually covered by the traditional anatomy spotter, whilst ensuring that the exam was closed book and limited collusion between students. Upon finishing the first round of examinations, it was considered between colleagues that this online assessment could be used in other, non-emergency, situations. To attempt to answer this, the discussion will be centred around the five parameters that make up the UTILTY formula outlined by Van Der Vleuten (1996); these being reliability, validity, educational impact, cost effectiveness, and acceptance.

4.1 | Reliability

To assess how reliably the exam was able to discriminate the best and worst students at spotter assessments, the online format was compared against traditional spotters, and a Pearson's correlation coefficient was calculated. This determined if students that performed well previously continued to perform well, and vice versa; poor performing students suffered the same fate in the online version. For the first round of assessments, year 2 students who sat LIFE218 and LIFE220 were compared against an average spotter score calculated from each student's performance in their four previous traditional spotters. Both coefficients were 0.49; these lower scores are likely due to the small number of questions asked in the first round of exams and the small sample size of students undergoing the online spotter. This issue with sample size is magnified further in the 19/20 cohort who sat only one traditional spotter; normally an easier, introductory test to help them become familiar with the exam format. When this was correlated against the online spotters undertook in first and second year the coefficients ranged from 0.31 to 0.42.

However, when the online spotters are compared against the student's overall anatomy score the coefficients suggest the online spotter is a reliable form of assessment. For the 18/19 cohort coefficients of 0.75 (LIFE218) and 0.66 (LIFE220) were recorded and for the 19/20 cohort all five coefficients were in the range of 0.65–0.75; Therefore, it can be assumed that the online spotters do reliably inform on the standard of the anatomist that is sitting them.

4.2 | Validity

This parameter considers whether the online spotter exam assesses the same skills and knowledge as a traditional spotter. The largest difference between the formats is the use of images compared to prosections. It is acknowledged that spatial understanding is a key area of assessment in spotters (Smith & McManus, 2015) and may be the only way to test for this skill (Ikah et al., 2015). It can be argued that anatomy is primarily about understanding the 3D structure and that it should be taught and assessed in such a context (Sagoo et al., 2016). In addition, some studies have reported that students who used images in computer-based spotters struggled to appreciate the relationship between structures (Choudhury et al., 2016), suggesting that images are not appropriate for this exam context.

However, spatial ability is the capability to 'perceive, retain, recognize or reproduce 3D objects in their current proportions when they are rotated in space, translated, juxtapositioned, projected, sectioned, re-assembled, inverted, re-orientated or verbally described' (Brenner et al., 2015). Indeed, all specimens, drawings, photographs and radiographs in spotter examinations are testing knowledge gained from 3D prosections that the students have previously studied (Vorstenbosch, Klaassen, Kooloos, et al., 2013). Thus, learning anatomy, by dissection or prosections, provides the students with the understanding of the structures in their 3D context (Daly, 2010). A test of spatial ability is, therefore, to be able to translate that knowledge to other forms, be that an image, a diagram, a medical scan, or a specimen that is orientated differently to how it may have been previously presented. Thus, it can be argued that despite the online spotter's use of 2D images, it still adequately tests a student's spatial awareness of anatomical structures, as students can connect prior 3D knowledge and use it to assess the context of each image. Students undertaking LIFE116 in 19/20 had access to prosections and had dissected up until March 2020. Students undertaking LIFE116 in 20/21 had very little exposure to human material throughout this period. Although the difference was not significant, there was a drop of 9% in LIFE116 mean scores between the two groups. Vorstenbosch, Klaassen, Donders, et al. (2013) have shown that learning anatomy increases spatial ability and, although tenuous, this 9% decrease could be explained by a lack of spatial ability development, given the reduced access to human material. For year 2 exams all students had a least two semesters of handling prosections.

Images in spotter examinations have been used previously with no evidence to suggest that anything other than a student's spatial knowledge of anatomy is being tested (Alraddadi et al., 2021; Brenner et al., 2015; Daly, 2010; Inuwa, Al Rawahy, et al., 2011; Meyer et al., 2016). However, in the Daly (2010) study the cohort could manipulate the images, which was not possible for this online spotter. Interestingly a study by Vorstenbosch, Klaassen, Kooloos, et al. (2013) did show increased difficulty when comparing photographic images to cross-sectional images suggesting that this does test extra abilities, nevertheless, cross-sectional medical imaging is something that was tested in both assessment forms and they are therefore directly comparable. Students' perception of the use of images has also been met with positive feedback (Inuwa, Taranikanti, et al., 2011; Meyer et al., 2016). A study that compared prosections with images during an objective structural practical examination saw a higher performance when students were examined on the cadaveric material. However, the authors raise the point that this is likely due to difficulties in interpreting the medical images rather than the 2D nature of the image (Sagoo et al., 2021).

The quality of images was brought up in the free text comments with six students commenting on the difficulty viewing them. Whether these were specific to the images themselves, the quality of the students Wi-Fi, the specification of the students hardware, or a general difficulty of using 2D images is unclear. Nevertheless, poor image quality will affect both the reliability and the validity of the exam and needs to be considered when forming the assessment.

A student's ability to orientate themselves when using images must also be considered. A magnified photograph with no quick reference points will likely take longer for the student to orientate than walking up to a specimen at a station. It could also lead to confusion if they orientate themselves wrongly. Whilst orientation of a specimen is a crucial ability for an anatomist or clinician to learn, and indeed it could be seen as advantageous to test for this to discriminate between students, it does constitute a disparity between the online and traditional spotter versions. Possible advances in the future could include a 'virtual classroom' whereby a video is recorded walking up to the specimen before zooming in and creating a freeze frame where the question will then appear. Simple orientation labels on the diagram would also help with this, however this would mark a difference from an in-person spotter as these specimens do not have orientation labels. Finally, the use of images will also prevent the student from using the tactile aspect of identifying anatomical structures (Smith & McManus, 2015). However, the in-person spotters do not permit touching of specimens and therefore the two modes are comparable.

Timing may have affected the validity of this examination, especially when considered with the above disparity in orientation. Seventy three percent of students thought that the time given for each question was fair, although four free text comments specifically mentioned the lack of time afforded to each question. Interestingly, Zhang et al. (2013) showed that there was no difference in mean marks between timed and untimed spotters and that timing should be a practical consideration rather than a cognitive one. However, it is likely there is a limit to this; while 30 s per question for LIFE116 is seemingly comparable with a traditional spotter; 30 s per part instead of a minute for two parts is different, and with the need for the students to orientate themselves first, may have been too little time. This may explain why only 58.5% of first year students felt that they had adequate time to answer the questions. If time is indeed no issue to student performance in a spotter exam it makes sense to increase this in any future tests.

The online spotter lacks handwritten answers which is a major difference when compared to the traditional method. Students were able to either verbalize their answers to the examiner, type their answer in the chat box, or do both. Numerous studies have mentioned that handwriting can be an issue in these exams, with examiners

struggling to understand what has been written (Choudhury et al., 2016; Shaibah & van der Vleuten, 2013). In some cases, tablets have been introduced in a traditional spotter to eliminate this source of possible unfairness (Polak et al., 2021). Losing marks due to poor handwriting is not a test of anatomical understanding or knowledge and therefore in this specific case the online spotter can be seen as a more valid test than a traditional handwritten spotter. The use of giving verbal answers must also be considered in future iterations of this exam. This was introduced to reduce anxiety amongst the students; by giving them options on how to answer they could tailor it to their preference and/or what their hardware would permit. However, verbal communications could lead to examiners mis-hearing responses especially if structures have similar sounding names and may even be open to unintentional biases from the examiners when recording what has been said.

4.3 | Cost

There are several practicalities to be considered when deciding whether it is viable to run an online spotter examination. Hosting an exam online saves laboratory time, which can instead be used for teaching (Daly, 2010). It also saves on technical time and costs, as the spotter examination does not need to be set up and taken down (Daly, 2010; Guimaraes et al., 2018). It prevents wear and tear of the cadaveric material, which saves prosecting time replacing the specimens (Guimaraes et al., 2018). This is an important consideration if the exam is early in the morning; often the specimens must be pinned the night before leading to excessive drying and possible degradation of the specimens. The use of prosections has additional drawbacks as the breadth of images available online will always outstrip that of an anatomy department. Therefore, the use of images will allow for questions and specimens to be changed year after year (Daly, 2010; Guimaraes et al., 2018), making it less likely that answers will be shared across cohorts (Meyer et al., 2016). It will also allow the introduction of variations and pathologies that may not exist in collections.

These are quite large savings in lab and technical time, and alongside other computer-based exams, which have been able to reduce academic staff time with automated marking and psychometric data generation (Guimaraes et al., 2018), the online spotter appears as an attractive alternative to the traditional format. However, for this online spotter, academic staff time is the largest cost that must be weighed against the above points. Staff time was extensive in formulating, designing, administrating, preparing students, and training staff members prior to the exam. On exam day it took the morning of nine academic members of staff to examine an average of four students each and following the exam it took the time of the module lead to mark answers and review the videos in cases of uncertainty. Whilst some of the preparatory time will be reduced if the format becomes established, this is still a large outlay that must be considered alongside the other parameters. In addition, it means this type of assessment can likely only be used on relatively small cohort groups. It is difficult to imagine a scenario where an institution has enough staff members to assess a large, 300 plus, undergraduate medical course. Adaptations could be explored; exams were carried out individually to prevent collusion, however a scenario could be envisaged whereby students take the exam in the same Zoom room and all type their answers. Care would have to be taken to ensure answers were delivered to the invigilator/proctor but this would decrease academic time considerably. Grouping the students would make the examination more in line with the Sadeesh et al. (2021) study and would help to make it more cost effective therefore opening up the format to larger cohorts.

4.4 | Educational impact

It is clear that examination drives learning (Ikah et al., 2015). Thus, it is believed that this online examination was successful in this aspect, as it is likely that students did not resort to learning anatomy purely theoretically when sent home during the COVID-19 pandemic. Instead, it increased the probability that students learnt anatomy by studying cadaveric photographs and images from atlases or online videos and not just standardized, diagrammatic versions. The danger with an online spotter examination in other, non-lockdown, cases is that students stay away from the anatomy laboratory and spend their revision studying from these same anatomical photographs. This has previously been overcome in computer-based spotter examinations by using either photographs of the students' dissection or prosections used in the lab, thus making it more likely that students would still learn from specimens in the lab (Daly, 2010; Inuwa et al., 2012). As discussed previously, this is crucial to understand the 3D spatial anatomy and the ability to apply it to 2D images. Using images of departmental material may of course be difficult with the various restrictions on the use of cadaveric images but is an interesting consideration to enhance the educational impact of an online spotter examination.

4.5 | Stakeholder acceptance

A statistically proven, reliable exam with constant monitoring to ensure validity can still falter if students, staff, or other stakeholders fundamentally dislike or are opposed to the form of assessment. Therefore, a questionnaire was devised to ascertain students' views on the online spotter and to feedback suggestions for improvement.

Reponses taken from the survey suggest that the students did find that this was an acceptable form of exam, with 82.5% of the students responding that it was a fair way to test their knowledge. Whilst there is no indication that students would rather take the online version, the results suggest that students may accept this form of assessment in a post-pandemic world in instances where it can be implemented. One free text comment did respond that *'I thought given the current circumstances the replacement online spotter was fair and very well organized'* suggesting that these responses may have been sympathetic to the situation and if implemented during normal times the responses may not have been as positive. This will have to be revaluated if this form is to run outside of a lockdown.

As mentioned previously students who have sat a similar computer-based exams have not been averse to the format (Inuwa, Taranikanti, et al., 2011). Those who preferred the computer-based format pointed to the quality of specimens online being better (and therefore easier to identify the structures), it being advantageous to be sat down in one place, and also, for this particular study, that there was an overall time limit (and not a limit for each question). It was noted, however, that the seniors preferred online compared to the juniors; likely because they were used to the test and because they had adapted their study to using atlases. This emphasizes the need to prepare students for this different format. This was a priority for us when designing the exam and, given the responses in the survey, was well received by the students.

Another computer-based spotter reported that 87% students felt comfortable with computers and two-thirds were happy with the image quality (Meyer et al., 2016) which is similar to results obtained in our questionnaire. However, a third of students in the Meyer et al. (2016) study disagreed with the statement that it was easy to orientate the images, which was not specifically commented upon in our survey, however it is still worth considering for future iterations of the online spotters.

Sadeesh et al. (2021) ran a similar online spotter on Zoom which used Google Forms for answers. It differed from our online spotter as 25 students were in a virtual waiting room as others were examined, our spotter avoided this to prevent cases of collusion. In this example 60% preferred traditional spotters, although only slightly more than half agreed that they could clearly identify the pointers and markers, which could explain the preference for traditional in-lab assessments. This assessment also had a viva voce element which is different from traditional spotters and therefore difficult to properly compare to our online assessment.

Lack of note taking has been cited as a shortcoming of computer based assessment (Inuwa et al., 2012) however, this spotter exam still allowed for that as students were not prevented from writing notes, or reminders to themselves during the process.

Student acceptance of this exam will also be driven by their own technical set ups. Poor image quality was mentioned by six students which, given the low number of students mentioning it, may indicate a problem with their hardware/software/internet connectivity as opposed to a poor image being placed on the exam. Internet connectivity has been mentioned in previous studies (Guimaraes et al., 2018; Inuwa, Taranikanti, et al., 2011) and this will always have to be taken into account, even the mere distraction of a student worrying about their internet connection may affect their test score, as one student did mention.

Free text comments suggested areas of improvements that were implemented in the second round of online spotters and some which can be included in any future assessments to improve student perception and therefore acceptance of the assessment. The most commented aspect was the shortness of the exam; in the first-round students were given 16 questions to answer. At that point, in the depths of the national lockdown, it was felt the shorter exam was still adequate in terms of sampling and carried less risk as there was less time for possible Wi-Fi outages, hardware breakages or staff or student interruptions. There was also a general desire to ease students' anxiety by 'getting it over with' quickly. However, students clearly identified that the importance of every question was therefore magnified, and this caused more concern. For the second round this was increased to 25 questions per exam.

Anecdotally, prior to the exams taking place, the largest source of anxiety from the student body was that they would be watched and have to respond directly to a member of staff. It was therefore surprising that less than 50% reported higher anxiety levels and only three text comments mentioned the awkwardness of being watched and 'judged' by the examiner. In addition, eight students said they felt less anxious and two used the word 'comfortable' because they were sitting the exam from home. Future exams could experiment with the use of cameras for both the student and the examiner to help reduce anxiety levels further.

5 | CONCLUSION

In conclusion, it is believed that this form of online spotter assessment, which was born out necessity in the COVID-19 pandemic is a viable form of assessment albeit in particular circumstances. Reliability and validity are comparable with traditional spotters and could be improved in future iterations of the exam. Student acceptance of this and other forms of online or computer-based spotters are positive, and again will likely improve as students get more accustomed to this. Whilst the laboratory usage and lack of exam set up are advantageous the need for numerous academic staff members to examine concurrently means that this format is only likely feasible for small cohorts of students. The implementation of the online spotter examination alongside written exams allowed for the assessment of all learning objectives during the COVID-19 pandemic. It prevented collusion amongst the student body and ensured that students continued to learn 'real', non-diagrammatic, anatomy. The adoption of this test in the future could allow the traditional spotter to keep up with modern, online learning.

ACKNOWLEDGMENTS

The authors acknowledge the support of all staff in the Human Anatomy Resource Centre at Liverpool, students on the BSc Anatomy and Human Biology degree at University of Liverpool, and Katherine Patrick for the suggestion of the title.

ORCID

Alistair P. Bond https://orcid.org/0000-0002-6160-0005
Thomas Butts https://orcid.org/0000-0002-1781-5287
Claire M. Tierney https://orcid.org/0000-0002-0079-9946

REFERENCES

Abrahams, P. H., Spratt, J. D., Loukas, M., & Van Schoor, A. N. (2020). *Abrahams' and McMinn's clinical atlas of human anatomy* (8th ed.). Elsevier Limited.

Alraddadi, A., Hoja, I., Alhawas, H., Khawaji, B., Alharbi, Y., Agha, S., Masuadi, E., & Magzoub, M. E. (2021). Introducing free response short answer questions in anatomy spot tests: Experiment study. *Surgical and Radiologic Anatomy*, *43*(4), 497–503. https://doi.org/10.1007/s00276-020-02550-3

Attardi, S. M., Harmon, D. J., Barremkala, M., Bentley, D. C., Brown, K. M., Dennis, J. F., Goldman, H. M., Harrell, K. M., Klein, B. A., Ramnanan, C. J., & Farkas, G. J. (2021). An analysis of anatomy education before and during Covid-19: August-December 2020. *Anatomical Sciences Education*, *15*, 5–26. https://doi.org/10.1002/ase.2152

Brassett, C., Cosker, T., Davies, D. C., Dockery, P., Gillingwater, T. H., Lee, T. C., Milz, S., Parson, S. H., Quondamatteo, F., & Wilkinson, T. (2020). COVID-19 and anatomy: Stimulus and initial response. *Journal of Anatomy*, *237*(3), 393–403. https://doi.org/10.1111/joa.13274

Brenner, E., Chirculescu, A. R. M., Reblet, C., & Smith, C. (2015). Assessment in anatomy. *Eur Journal of Anatomy*, *19*(1), 105–124.

Choudhury, B., Gouldsborough, I., & Shaw, F. L. (2016). The intelligent anatomy spotter: A new approach to incorporate higher levels of Bloom's taxonomy. *Anatomical Sciences Education*, *9*(5), 440–445. https://doi.org/10.1002/ase.1588

Daly, F. J. (2010). Use of electronic anatomy practical examinations for remediating "At Risk" students. *Anatomical Sciences Education*, *3*(1), 46–49. https://doi.org/10.1002/ase.120

Drake, R. L., Vogl, A. W., & Mitchell, A. W. M. (2020). *Gray's anatomy for students* (4th ed.). Elsevier.

Guimaraes, B., Ribeiro, J., Cruz, B., Ferreira, A., Alves, H., Cruz-Correia, R., Madeira, M. D., & Ferreira, M. A. (2018). Performance equivalency between computer-based and traditional pen-and-paper assessment: A case study in clinical anatomy. *Anatomical Sciences Education*, *11*(2), 124–136. https://doi.org/10.1002/ase.1720

Harmon, D. J., Attardi, S. M., Barremkala, M., Bentley, D. C., Brown, K. M., Dennis, J. F., Goldman, H. M., Harrell, K. M., Klein, B. A., Ramnanan, C. J., Richtsmeier, J. T., & Farkas, G. J. (2021). An analysis of anatomy education before and during Covid-19: May-August 2020. *Anatomical Sciences Education*, *14*(2), 132–147. https://doi.org/10.1002/ase.2051

Ikah, D. S. K., Finn, G. M., Swamy, M., White, P. M., & McLachlan, J. C. (2015). Clinical vignettes improve performance in anatomy practical assessment. *Anatomical Sciences Education*, *8*(3), 221–229. https://doi.org/10.1002/ase.1471

Inuwa, I. M., Al Rawahy, M., Taranikanti, V., & Habbal, O. (2011). Anatomy "steeplechase" online: Necessity sometimes is the catalyst for innovation. *Anatomical Sciences Education*, *4*(2), 115–118. https://doi.org/10.1002/ase.188

Inuwa, I. M., Taranikanti, V., Al-Rawahy, M., & Habbal, O. (2011). Perceptions and attitudes of medical students towards two methods of assessing practical anatomy knowledge. *Sultan Qaboos University Medical Journal*, *11*(3), 383–390.

Inuwa, I. M., Taranikanti, V., Al-Rawahy, M., & Habbal, O. (2012). Anatomy practical examinations: How does student performance on computerized evaluation compare with the traditional format? *Anatomical Sciences Education*, *5*(1), 27–32. https://doi.org/10.1002/ase.254

Merzougui, W. H., Myers, M. A., Hall, S., Elmansouri, A., Parker, R., Robson, A. D., Kurn, O., Parrott, R., Geoghegan, K., Harrison, C. H., Anbu, D., Dean, O., & Border, S. (2021). Multiple-choice versus open-ended questions in advanced clinical neuroanatomy: Using a National Neuroanatomy Assessment to investigate variability in performance using different question types. *Anatomical Sciences Education*, *14*(3), 296–305. https://doi.org/10.1002/ase.2053

Meyer, A. J., Innes, S. I., Stomski, N. J., & Armson, A. J. (2016). Student performance on practical gross anatomy examinations is not affected by assessment modality. *Anatomical Sciences Education*, *9*(2), 111–120. https://doi.org/10.1002/ase.1542

Polak, K., Gielecki, J. S., & Zurada, A. (2021). Tablet-assisted objective structured spotter practical examination (TOSSPE): Advantages of an innovative anatomy spotter examination method for medical student assessment. *Anatomical Sciences Education*, *15*, 1060–1073. https://doi.org/10.1002/ase.2131

Rowland, S., Ahmed, K., Davies, D. C., Ashrafian, H., Patel, V., Darzi, A., Paraskeva, P. A., & Athanasiou, T. (2011). Assessment of anatomical knowledge for clinical practice: Perceptions of clinicians and students. *Surgical and Radiologic Anatomy*, *33*(3), 263–269. https://doi.org/10.1007/s00276-010-0748-8

Sadeesh, T., Prabavathy, G., & Ganapathy, A. (2021). Evaluation of undergraduate medical students' preference to human anatomy practical assessment methodology: A comparison between online and traditional methods. *Surgical and Radiologic Anatomy*, *43*(4), 531–535. https://doi.org/10.1007/s00276-020-02637-x

Sagoo, M. G., Smith, C. F., & Gosden, E. (2016). Assessment of anatomical knowledge by practical examinations: The effect of question design on student performance. *Anatomical Sciences Education*, *9*(5), 446–452. https://doi.org/10.1002/ase.1597

Sagoo, M. G., Vorstenbosch, M. A. T. M., Bazira, P. J., Ellis, H., Kambouri, M., & Owen, C. (2021). Online assessment of applied anatomy knowledge: The effect of images on medical Students' performance. *Anatomical Sciences Education*, *14*(3), 342–351. https://doi.org/10.1002/ase.1965

Schubert, S., Schnabel, K. P., & Winkelmann, A. (2009). Assessment of spatial anatomical knowledge with a 'three-dimensional multiple choice test' (3D-MC). *Medical Teacher*, *31*(1), E13–E17. https://doi.org/10.1080/01421590802334325

Shaibah, H. S., & van der Vleuten, C. P. M. (2013). The validity of multiple choice practical examinations as an alternative to traditional free response examination formats in gross anatomy. *Anatomical Sciences Education*, *6*(3), 149–156. https://doi.org/10.1002/ase.1325

Smith, C. F., & McManus, B. (2015). The integrated anatomy practical paper: A robust assessment method for anatomy education today. *Anatomical Sciences Education*, *8*(1), 63–73. https://doi.org/10.1002/ase.1454

Van Der Vleuten, C. P. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education: Theory and Practice*, *1*(1), 41–67. https://doi.org/10.1007/BF00596229

Vorstenbosch, M. A., Klaassen, T. P., Donders, A. R., Kooloos, J. G., Bolhuis, S. M., & Laan, R. F. (2013). Learning anatomy enhances spatial ability. *Anatomical Sciences Education*, *6*(4), 257–262. https://doi.org/10.1002/ase.1346

Vorstenbosch, M. A. T. M., Klaassen, T. P. F. M., Kooloos, J. G. M., Bolhuis, S. M., & Laan, R. F. J. M. (2013). Do images influence assessment in anatomy? Exploring the effect of images on item difficulty and item discrimination. *Anatomical Sciences Education*, *6*(1), 29–41. https://doi.org/10.1002/ase.1290

Zhang, G., Fenderson, B. A., Schmidt, R. R., & Veloski, J. J. (2013). Equivalence of students' scores on timed and untimed anatomy practical examinations. *Anatomical Sciences Education*, *6*(5), 281–285. https://doi.org/10.1002/ase.1357

Zoom Video Communications. (2022). *Zoom*. [Computer program] Available at https://zoom.us (Accessed 1 March 2022).

How to cite this article: Bond, A. P., Butts, T., & Tierney, C. M. (2023). Spot(ters) the difference: Bringing traditional anatomical examinations online. *Clinical Anatomy*, 1–10. https://doi.org/10.1002/ca.24092