Continuous Controlled Conditions Assessment to Promote Frequent Revision Engagement

Chris Moore
University of the West of England, Bristol

With assessment outcomes being the primary driving force for students, revisiting learning material usually occurs only during the lead up to end of learning time point assessment. An alternative assessment was developed to promote continuous engagement with material by introducing, online, controlled conditions, monthly assessment, including re-testing of material from earlier exams, on a large cohort first year module. Average cohort results showed some improvement in overall assessment outcomes when compared to previous cohorts using traditional methods. Transition to methods of assessment that encourage continuous engagement with material, fosters an improved attitude to learning and enables in-year.

INTRODUCTION

Traditional forms of controlled conditions assessment on courses in the scientific field are designed based on the model of semesterised or end of year time points that assess the student's knowledge and understanding of the course material. In first and second years of undergraduate study these often include a proportion of multiple choice questions (typically best *single* answer) completed using OMR sheets, as well as perhaps short answer (1-2 mark) and medium answer (5-10 mark) free text based questions in order to combine assessment methods and move beyond the superficial, surface learning sometimes professed to be a disadvantage of MCQs (Gibbs, 2010; Brady, 2005) . Essay length questions are sometimes included at level one, depending on the subject matter, more so at level two, and almost exclusively at level three as the entire basis of a final year exam.

These forms of questions all have their merits, and essays are, arguably, the best way of assessing a student's ability to critically evaluate academic content, particularly beyond that taught as part of the core curriculum. At earlier stages though, especially first year, a question format that ultimately provides evidence of having understood the material such that one can be confident in moving onto application of in year two, is usually more appropriate. It is also important to consider that often at year one, the skills necessary for essay writing under controlled conditions have not been taught/practised, and where establishing a foundation level of knowledge and understanding is perhaps more important in helping students develop the means to apply that knowledge.

There is also the issue of large class sizes in years one and often two at university level. Use of an assessment model more in depth and more time consuming in the completion and grading is often precluded by the number of students one must assess (Fortes & Tchantchane, 2010). Light and Cox (2001) argue that any assessment in order to be acceptable to faculty members must not unduly add to an academic's already excessive workload. The trend of assessors therefore, is more often than not one that can assess a large number of students on a wide range of content in a relatively short time with minimal marking required (Brady, 2005).

Whatever the level and whatever the question format however, these models all have one issue in common – timely engagement with the content. An end of year or end of semester exam only requires that the student taking it engage with/revise their material/notes a short time beforehand in order to pass; usually over Christmas or Easter break. Though they are of course engaging with the material at that time and might grade well, there is no guarantee that the information remains with them even shortly after the exam. This

also encourages the mind set of 'reading to pass,' not 'reading to learn,' not to mention exacerbating the current problems we face with student attendance (Wilkinson, 2017; Meyer *et al,* 2016; Gump, 2005). With the knowledge in advance that all course material will be made available prior to the exams whether one attended or not, there is little incentive for potentially 'at risk' students to attend in the first place, let alone engage with the lecture/seminar should they actually attend.

The flipped classroom model of learning; that of providing material to engage with before attending the timetabled session in which discussion becomes the learning exchange over didactic assimilation, has proven popular in recent years. This strategy to improve student understanding of subject matter through active learning provides more flexibility for educators, and thus adds greater variety for the students that ideally results in improved attendance and engagement (Schmidt and Ralph, 2016). It is believed that greater responsibility for one's own learning by way of preparation increases the learner's engagement (Fulton, 2012). This approach goes some way in addressing student engagement, though is not without its disadvantages, particularly in a large cohort setting where smaller discussions within the class could be problematic in terms of certainty of widescale engagement. There have however, been successes reported in a review by O'Flaherty and Phillips (2015), where large first year foundational STEM courses showed great willingness to adopt the model. It has not been found though, whether the assessment model used for these flipped classroom units of study, at any level of education, was anything other than the traditional end of year or end of semester format. This therefore may still mean that whatever the model of teaching used, revision remains a last minute endeavour and thus does not demonstrate whether the flipped classroom improves long term retention and use of knowledge gained beyond that end of unit assessment.

In an effort to address this issue, new technology was used to reshape the current level one MCQ and semesterised based exams on the module *Anatomy and Physiology*. This involved turning two, single semester based, 90 minute/90 question paper based exams into six, monthly, online and automatically graded exams of 30 minutes each. These exams were sat at the end of six practical classes across the academic year, and the students final grade was based on their best five of the six. The aim of this redesign was to encourage students to engage with their notes *all year round*, even if just frequently glancing, to discourage cramming, to improve attendance, and to promote engagement in class.

In addition, though controlled conditions summative assessment in its traditional guise provides a measurement of the learning of the student, it does not normally provide an opportunity to learn *from* the assessment; the so-coined 'assessment *for* learning' as opposed to simply an 'assessment *of* learning'. The lessons of how to improve and which pitfalls to avoid fall to that of the coursework based assignment where feedback is normally accessible and available to the student after it has been marked. But arguably any assignment, controlled conditions or otherwise, should offer actionable feedback that might aid students with similar assignments in the future, or even simply to improve performance in resit exams. Formative assessment of course provides this, and practice exam papers can be useful. But neither guarantee or even encourage active and total participation as the process does not include an assigned grade that counts towards a final qualification; the concept of 'giving it your all' compared to a summative assessment.

One of the major benefits of continuous assessment however, particularly one delivered and graded online, is that of the ability to reflect on the most recent grade and act accordingly by developing new strategies for learning and revision in order to perform better next time. Therefore a mechanism of simple feedback provision was introduced to each monthly exam that had not previously been offered in any paper based MCQ summative assessments on the module.

METHODS

A first year class of undergraduate students studying Biomedical, Biological, or Healthcare sciences degree programmes at Bristol UWE undertook a series of timetabled monthly exams across the academic years 2016/17 and 2017/18 on the first year module *Anatomy and Physiology*. This format of controlled conditions assessment was used in place of the model used in previous years; that of two end of semester exams carried out as 90 single best answer Multiple Choice Questions (MCQ) each based on content from the respective semester's lectures. Six of these monthly exams were sat across the year on a monthly basis,

where the marks from the five highest scoring would go towards the final summative assessment grade for the module.

Each monthly exam lasted thirty minutes and was comprised of 30 questions. Those 30 questions were split into two parts, both completed in the same sitting. Two thirds of the questions assessed lecture content from the previous two lecture sessions (each session being three hours in duration). The final third assessed lecture content from all previous lectures that year so far, with the 10 questions within that third drawn randomly from a bank that increased in size as each lecture was added to it following initial assessment as part of the two thirds new content. This model effectively duplicated the number of questions and total time of assessment as previous paper based and semesterised exams.

Exams were carried out in class at the end of the laboratory session during the practical week that followed on from lecture week. Laboratory sessions consisted of approximately 50 students, necessitating five runs of each exam across the week. Normal controlled conditions assessment regulations applied. Each student was given a 10 inch *Learnpad* tablet to complete their exam. These tablets, designed specifically for teaching and learning, enable certain content, links or apps to be sent remotely to any selected tablets or groups of tablets, and prevent access to any content or other internet use beyond the selected content. In this case students only had access to a link that sent them directly to the site within Bristol UWE's Virtual Learning Environment (VLE), *Blackboard* where the exam was hosted, which prompted each student to log into Blackboard using their Bristol UWE student login.

Exams were created using *ispring quizmaker,* and published online as SCORM packages, hosted within Blackboard and that automatically graded each submission, and sent the results to the VLE grade centre. On opening the exam, students had to enter their name and Bristol UWE e-mail address, which would then enable a copy of their submitted responses to be sent to an e-mail account as a back-up and for reference. The question format of the exams included MCQ, but also multiple choice MCQ, word bank, matching pairs, hotspot, sequential, and other *Varied Online Question* (VOQ) formats. The order of these questions as well as the order of answer options within each question was randomised. Each twenty question new material section was randomly drawn from a larger bank of questions within the relevant topic to prevent students in different sittings in that week from being able to pass on information regarding questions included, order of questions, and potential answer option orders. Each sitting of each exam also required a different password to access it, which was only given verbally once students were ready to begin to ensure only those present could complete the exam. Of the six exams among the twelve timetabled practical sessions across the year, the highest scoring five of each student counted towards their final grade.

Students completed each exam sequentially, but were able to flag questions they wished to return to later, and to select which question they wished to attempt from a drop down menu. Though answers could be entered and options selected, submission of final answers was not until the student clicked 'submit all' when satisfied with all of their responses or when the 30 minutes was up, whichever was sooner. On submission of all responses, students were able to review their answers to identify areas of strength and weakness. This feedback was provided with a simple right/wrong and with reveal of the correct or acceptable answers. Students were only permitted one attempt of each exam.

Data was collected from the grade centre within Blackboard from across the year to analyse cohort performance as the year progressed by looking at both outcomes from exam to exam as well as at final exam outcomes, and compared to previous years in terms of overall exam outcomes.

RESULTS

The final exam cohort averages for the academic years 2013/14, 14/15, 15/16, 16/17, and 17/18 were 43%, 53%, 53%, 44%, and 51% respectively. In terms of performance as the academic year went on, average grade across each cohort of the two years this form of assessment was used reduced from exam one to exam six. Average cohort grade for each exam in the 2017/18 cohort compared to the 2016/17 cohort however, was higher in each exam.

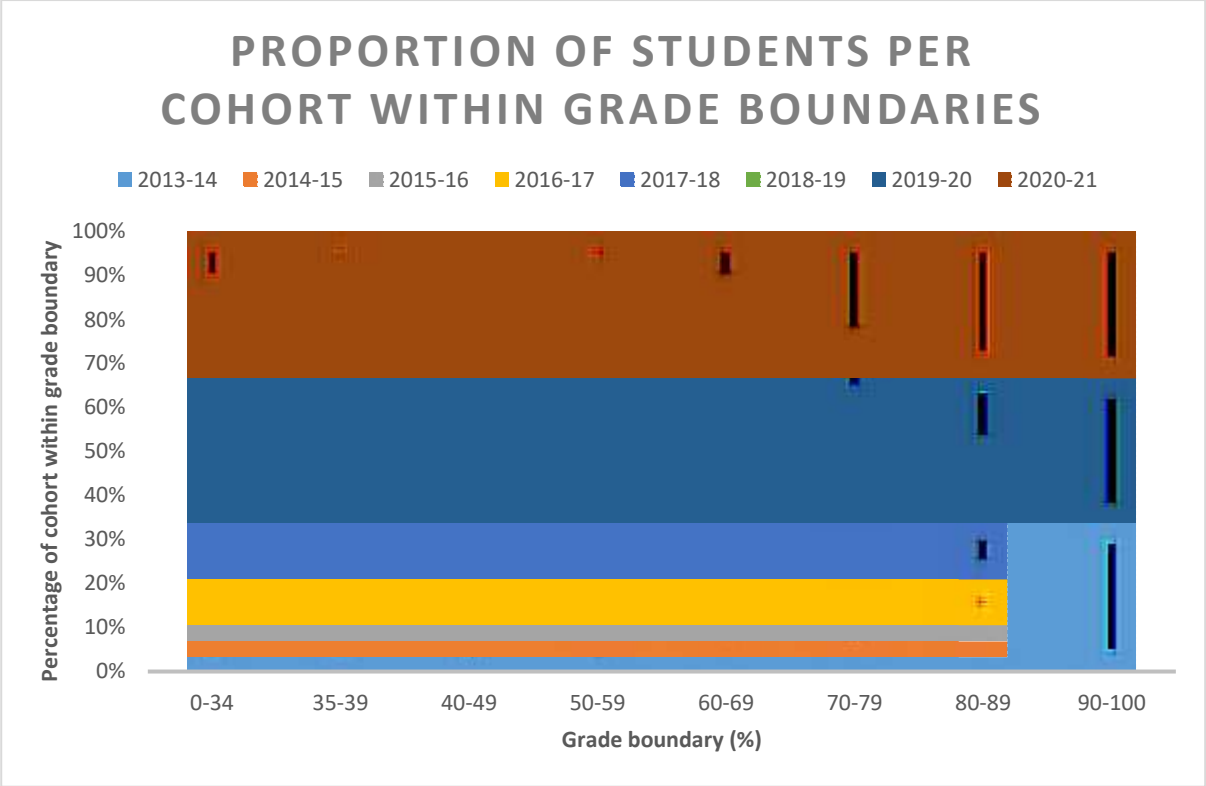
TABLE 1: CLASS AVERAGES FOR EACH OF THE SIX EXAMS ACROSS THE ACADEMIC YEARS 2016-2021

	Average percentage grade					
	Exam 1	Exam 2	Exam 3	Exam 4	Exam 5	Exam 6
All Students (n=231) Academic Year 2016-17	54	48	42	42	43	36
All Students (n=235) Academic Year 2017-18	57	53	57	55	45	40
All Students (n=235) Academic Year 2018-19	57	54	55	61	48	46
All Students (n=239) Academic Year 2019-20	58	56	61	63	57	51
All Students (n=198) Academic Year 2020-21	69	58	60	63	61	58

Average end of year exam grade for traditional exams was 50%, while that of monthly exams across the range was 55%.

In contrast to traditional MCQ based assessment in academic years 2013/14 to 2015/16, with only some exceptions, the proportion of students within higher grade boundaries at the end of the academic year was generally higher for those who sat the monthly exam form of summative assessment. Of note in 2013/14 is the significantly higher proportion of students achieving grades between 0-35%; a fail requiring a resit. This improves considerably over the next two years but then increases again once monthly exams are introduced, though not to the same degree as 2013/14. And higher grade boundary performance improves considerably again in those cohorts sitting monthly exams over end of semester exams. With the exception of 2013/14 however, student achievement of final grades does not vary significantly in those middle grade boundaries. But again, even within those boundaries, the 2017/18 cohort undertaking monthly exams do perform, on average, better than previous years.

FIGURE 1: PROPORTION OF EACH ACADEMIC YEAR COHORT ACHIEVING FINAL MODULE EXAM GRADES WITHIN GRADE BOUNDARIES FOR FIVE COPNSECTUTIVE YEARS OF STUDENTS. YEARS 2013/14 - 2015/16 BASED ON TWO END OF SEMESTER PAPER MCQS EXAMS. YEARS 2016/17 ONWARDS ARE ONLINE MONTHLY EXAMS



Results for exam 6 in those years of monthly assessments decreases on average due to a reduced number of students sitting that final exam.

DISCUSSION

Frequent testing is not something new to the educational platform, with pop quizzes remaining a popular method of feedback and reflective assessment in many U.S. educational institutions (Cicirello, 2009). Or even to higher education in the UK and Europe where several works have investigated student perceptions of continuous summative assessment. Studies such as Hernandez (2012) on humanities courses in Ireland, and Trotter (2006) who found the over-riding impression from an analysis of student interviews and the questionnaires following use of continuous, summative assessment was that students were extremely positive. However, an issue that is likely partly responsible for limited uptake of summative assessments at more than one or two major time points is that of administration and marking workload associated with additional assessment phases. There is the impact on student performance as a major metric of institutional success to consider of course. Despite the positive student attitudes to continuous summative assessment, Trotter (2006) also concluded that although continuous summative assessment does carry with it, significant administrative burden, the effects of an enhanced student experience should in fact outweigh staff burden. Contradictory to the beliefs of Light and Cox (2001) on practicable and realistic assessment models, this statement is in part correct. Academia should always look to improve student outcomes and experiences without compromising on robustness of their education. But stress and workload demands of academic staff remain a serious concern throughout the sector (Darabi, Macaskill and Reidy, 2017; Graham, 2015; Kenny, 2015), and quality of feedback may suffer as a result of increasing this demand under the guise of student experience outweighing staff burden. Therefore the key to successful uptake and implementation of any continuous assessment would seem to be a balance of maximum benefit to students and minimal burden on staff. The present study achieves this by creating an online platform that provides varied and robust questions formats beyond the simple MCQ that themselves have strengths and

weaknesses in relation to their robustness (Donnelly, 2014), and automates the marking and feedback process. Although there remains an initial outlay in effort by staff in the setup of this form of summative assessment, it is no more so than generating the same number of questions as an end of semester or end of year examination.

The use of technology is also a factor to consider in moving to newer models of assessment. The automated marking and control aspects certainly improve scalability and variation in the form of questions being posed, but it is not without its own potential pitfalls. Chief among these is the requirement of the online quiz resource to communicate frequently with the VLE, Blackboard. The technology used in this study, the *learnpad* tablets, had to ping the VLE with every submitted answer so that the cumulative score could be recorded and saved. In order to ensure no data was lost, the devices attempted to ping the VLE every 10,000 milliseconds in addition to each answer submission. This required a stable Wi-Fi connection, and there were multiple instances early on where users experienced drops in this communication, resulting in closure of the exam following repeated failed attempts of the *learnpads* to receive a reciprocal ping from the VLE. The problem was identified as excessive consumption of bandwidth, which was traced back to student's personal devices remaining on during the exam despite instructions to the contrary, and picking up personal e-mails and social media messages. Reinforcing the importance of this issue to the students as the result could be the loss of their own submission, quickly resolved the issue. This did not impact any student submissions beyond a slight delay in reopening the exam link however, as all exams were designed to resume where left off and therefore all prior answer submissions remained in place in the event a crash occurred. But this does highlight the importance of student adherence to controlled conditions assessment environments and rules, even when not undertaking an exam in a traditional setting. Though perhaps more importantly, significant investment in the IT infrastructure of institutions committed to moving to a more digital form of assessment, whatever the model used (Spurlin, 2007).

The use of *learnpads* in conjunction with the *ispring* software proved to be a hugely beneficial blending of technology in terms of the administering and success of this model. Though automated marking removes much of the freedom of free text based answers, the use of questions necessitating simple factual recall of terms that could not easily be reinterpreted did allow for some questions where the answer could not so easily be discerned from the question. This is an inherent problem with many MCQ style questions, where correct deduction of the answer from those presented may not always be the most efficient method of student demonstration of actual knowledge. Further, the visual nature of many question formats within the *ispring* package encourages improved inclusivity for many types of learners. In particular, the question format "hot spot" provides a more authentic means of assessing knowledge of anatomy by requiring the student to identify a particular structure within a larger image of a system/organ/tissue. Students assessed using this style of question therefore improve their understanding of individual anatomical structures by contextualising them when they may originally have been taught *in isolate*.

Feedback too, proved a useful additional measure. Students, on completion of their monthly exam, could review their answers and be shown, question by question, what the correct answer should have been. Though they were not permitted to record this feedback in any way and could not access that review once closing the exam and leaving the room, they did gain an appreciation of which areas or topics they had performed strongest or weakest in and therefore judge more accurately, where to focus their revision efforts. This enabled a process of continuous improvement, reducing anxiety by giving students the chance to do better next time, rather than rely on a single assessment at the end of the year to determine their module outcome.

Primary considerations as to the validity of this approach to controlled conditions assessment are those of perceived relationships between engagement with learning materials and assessment outcomes, uptake of the idea and mechanism of assessment by all students participating in the study, and cohort variation year by year irrespective of the method of assessment practice. The data shows that, in comparison to traditional end of year assessments, there is no significant improvement in assessment outcome. The model presented here however, is built on a simple fundamental that is beyond the control of the way in which the assessments themselves are conducted; engagement. Success of the model is only possible if the manner in which the students involved approach their lectures and their revision differs from that of end of

semester/end of year assessments. If students assessed in this way continue only to revisit and re-read their material on the weekend before the week of the monthly exam, then long term retention of the information and building on its understanding remains elusive. Their grades obtained become comparable to single time point assessments, and thus the benefits of continuous assessment and reflecting on past performances become moot.

Indeed, in the results from the two years in which students undertook monthly exams instead of semesterised exams, though the most recent year did better on average in each monthly exam, the performance from exam to exam in each year declines. One could argue that this is due to the increase in total material requiring revision as the year progresses, with only two topics to study and be examined on in the first monthly exam, but twelve topics to be completely familiar with by the time the sixth monthly exam is sat; although the number of questions on that larger body of revisited material gets fewer with each successive exam. Individual student performance within these averages is likely a more compounding factor however. Those students whose level of engagement fell as the year went on may have affected the cohort data. And those who had performed poorly earlier in the year would have needed to sit their sixth exam to boost their average of their best five. Whereas those who had done well in the first five exams may not have needed to sit the final exam and thus may have elected to be absent from the last practical due to it being the end of the academic year and revision for upcoming end of year exams on other modules taking precedence. This could skew the overall data by a greater proportion of the takers of the final exam being those who performed worse earlier in the year. In support of this possibility, attendance in the last practical/exam was noted to be less compared to the rest of the year, with a decrease in total number of students with a result for exam six. This potential effect of engagement was realised part way through the second year of conducting exams in this way, where inquiries with students across several practical classes in one particular week revealed a mixed attitude to continuous revision in order to benefit from continuous assessment. All students had, on several occasions throughout the academic year, been advised about the need to continuously revisit their notes and lecture material, even if only briefly, to maximise their chances of success. But uptake and consideration of this message will always differ from student to student, for a variety of reasons. Nonetheless, subsequent study of some individual grades and performance from exam to exam following these inquiries revealed that many of those who claimed to frequently review their material did in fact perform better than those who admitted to having only revised on the weekend prior to their monthly exam. However, because this post-testing closer look was based on anecdotal information asked ad hoc, it is not possible to provide the data here or conduct any deeper analysis. It is intended that a survey will be distributed to students in an upcoming cohort to more accurately and completely gather this data while retaining anonymity in regards survey responses mapping to individual assessment outcomes.

Further work following on from this study proposes that the same cohorts evaluated using this method are assessed on the same material at the start of their second year of study on the appropriate module that builds on learning outcomes from the first year of study. Though a not insignificant period of time will have passed since these students likely last engaged with their university learning (the summer break), value in wider use of this method of continuous assessment as a means of promoting greater and more frequent engagement with the learning material could be proven through how well second year students perform in a "pre-second year/post-first year" formative assessment to gauge readiness for the second year of study.

REFERENCES

Brady, A. M. (2005). Assessment of learning with multiple-choice questions. *Nurse Education in Practice*, *5*(4), 238-242.

Cicirello, V. A. (2009). On the role and effectiveness of pop quizzes in CS1. *ACM SIGCSE Bulletin*, *41*(1), 286-290.

Darabi, M., Macaskill, A., & Reidy, L. (2017). A qualitative study of the UK academic role: positive features, negative aspects and associated stressors in a mainly teaching-focused university. *Journal of Further and Higher Education*, *41*(4), 566-580.

Donnelly, C. (2014). The use of case based multiple choice questions for assessing large group teaching: implications on student's learning. *Irish Journal of Academic Practice*, *3*(1), 12.

Fortes, P. C., & Tchantchane, A. (2010). Dealing with large classes: A real challenge. *Procedia-Social and Behavioral Sciences*, *8*, 272-280.

Fulton, K. (2012). The flipped classroom: Transforming education at Byron high school. *The Journal*, *39*(3), 18-20.

Gibbs, G. (2010). *Using assessment to support student learning*. Leeds Met Press.

Graham, A. T. (2015). Academic staff performance and workload in higher education in the UK: the conceptual dichotomy. *Journal of Further and Higher Education*, *39*(5), 665-679.

Gump, S. E. (2005). The cost of cutting class: Attendance as a predictor of success. *College Teaching*, *53*(1), 21-26.

Hernández, R. (2012). Does continuous assessment in higher education support student learning?. *Higher education*, *64*(4), 489-502.

Kenny, A. (2015). Work-related Stress: Survey of academic staff in the Institutes of Technology Sector. *Kenny A.(2015) Work-related Stress: Survey of academic staff in the Institutes of Technology Sector, DIT Level*, *3*.

Light, G., & Cox, R. (2001). Assessing: student assessment. *Learning and Teaching in Higher Education: The Reflective Practitioner. London: Paul Chapman Publishing*.

Meyer, E. R., Notebaert, A. J., Perkins, E., Lynch, J. C., Simpson, K. L., & Darling, R. D. (2016). Does lecture attendance matter? A study of exam performance in medical neurobiology. *The FASEB Journal*, *30*, 569-10.

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The internet and higher education*, *25*, 85-95.

Schmidt, S. M., & Ralph, D. L. (2016). The flipped classroom: A twist on teaching. *Contemporary Issues in Education Research (CIER)*, *9*(1), 1-6.

Spurlin, J. E. (2007). Using needs assessment as a holistic means for improving technology Infrastructure. *ELI Paper*, *3*, 1-10.

Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education*, *31*(5), 505-521.

MAILING INFORMATION

Dr Chris Moore
Dept Applied Sciences, UWE Bristol, Frenchay Campus, Coldharbour lane, Bristol, BS16 1QY
Chris3.moore@uwe.ac.uk