

This is a pre-print of “More than a Lucky Guess? Confidence-Based Multiple-Choice Questions for Worried Law Students”, Schmid Evelyne dans Astleitner Hermann, Deibl Ines, Lagodny Otto, Loos Claus, Warty Patrick, Zumbach Jörg (eds.) Tagungsband zur 2. Fachtagung Rechtsdidaktik, Nomos, 2019. <https://www.nomos-shop.de/Astleitner-Deibl-Lagodny-Warty-Zumbach-Rechtsdidaktik-zwischen-Theorie-Praxis/productview.aspx?product=39981>

# More than a Lucky Guess? Confidence-Based Multiple-Choice Questions for Worried Law Students

## *1. Introduction and Context*

It matters whether a law professional can aptly estimate his or her own skills and competences. A good lawyer or jurist knows when he or she needs to undertake further research and does so to the extent necessary to enable him or herself to come up with a solid answer. If he or she knows the answer to a problem – great. But if the legal professional is convinced that he or she knows the correct answer but this answer is actually wrong, this mistake can have serious negative consequences. Vice versa, an under-confident but actually competent professional who underestimates his or her own skills can lead to unnecessary costs to a client, overly long research or other opportunity costs. The ability to adequately self-assess one's own level of competence is thus an important professional and academic quality. Yet, building students' ability to aptly self-estimate their competence is rarely explicitly stated as an objective in higher education.

It also matters how students assess the level of difficulty of a particular course or subject matter and what interferences they draw for their own learning. It is well-known that the learning context can greatly influence students' motivation and engagement with a subject matter.<sup>1</sup> If students approach a new class with low confidence and worry that a subject matter is inherently unintelligible, what's the point of investing time in attending lectures, revising notes or studying with a textbook? Vice versa, if students believe that they can significantly influence their grade at the exam with the right type and amount of investment, they are more likely to engage in goal-oriented learning.<sup>2</sup> For reasons further stated below, I was interested in a teaching intervention that would give students an opportunity to obtain feedback both on their learning as well as on their self-assessment skills while at the same time building their confidence and easing their worries towards a new subject matter. Teaching relatively large groups, I also have a practical interest in multiple-choice questions (MCQ) and wanted to critically study their potential for assessing large cohorts in formative and summative settings. Yet, I was

<sup>1</sup> *Biggs*, *Teaching for Quality Learning at University: What the Student Does*, Philadelphia 2011, p. 39 (39-45).

<sup>2</sup> *Gibbs*, *How Assessment Frames Student Learning*, in: Bryan/Clegg (ed.), *Innovative Assessment in Higher Education*, London/New York 2006, p. 23, (stating that assessments »frame learning, create learning activity and orient all aspects of learning behavior«).

### *More than a Lucky Guess?*

conscious of the widespread reticence towards MCQ as well as the fact that traditional multiple-choice exams fail to distinguish *confident* correct answers from correct answers that are correct, but simply the result of a lucky guess. In short, I wanted to try a new method and gather some empirical evidence to inform my own future teaching practice. When law teachers endeavour to do this, a challenge is, of course, that undertaking empirical research is usually not the key business of law teachers and something about which most legal scholars have only limited methodological knowledge. Conducting the research around this teaching intervention thus came with the risk of »dabbling in other people's fields«. <sup>3</sup> Yet, despite the limitations and shortcomings of what I hoped to be a »good enough« approach, the empirical evidence can at least provide a source for inspiration. Moreover, the need for a practice of scholarship of teaching and learning is undisputed and evaluating one's teaching interventions is clearly encouraged (if not friendly imposed) on new teachers in UK universities. <sup>4</sup>

In any event, I decided to use confidence-based multiple choice tests in my European Union (EU) Law class. Confidence-based MCQ quizzes or tests not only ask students to provide the correct answer(s) but also asks them to rate their own answers according to how confident they feel about their responses. The underlying idea of this approach is that a lawyer – just like a medical doctor – not only needs to possess information about the law but he or she should also be able to assess his or her own certitude in giving, for instance, a specific piece of advice to clients, when critically analysing a court judgment or reviewing legal literature. The purpose of this chapter is to reflect on my experience of testing confident-based MCQ in an undergraduate law course. The remainder of part I will provide information about the context of the teaching intervention and explain what has been done and how the exercise was evaluated. Part II presents the empirical data. Part III contains the discussion and outlines the rationale for confidence-based MCQ and engages with the main concern that can be raised against using this approach. Part IV

<sup>3</sup> In his conference contribution, Patrick Wartho emphatically called for a practice of scholarship of teaching and learning and overcoming the challenges of dabbling in other fields. On the particular challenges of not (or not fully) mastering the necessary methods skills, see *Huber*, *Scholarship of Teaching and Learning: Konzept, Geschichte, Formen, Entwicklungsaufgaben*, in: Huber et al. (ed.), *Forschendes Lehren Im Eigenen Fach: Scholarship of Teaching and Learning in Beispielen*, Bielefeld 2014, p. 19 (30-32).

<sup>4</sup> As a new member of teaching staff, my contract contained a probatory requirement to obtain the Postgraduate Certificate in Teaching and Learning in Higher Education, a British teaching qualification for university lecturers leading to formal professional recognition with the UK Higher Education Academy. The research for this contribution was part of my portfolio for this programme.

concludes by reflecting on the implications for future teaching practice.

### *1. Context of the Intervention*

A few years ago, I was teaching EU Law for undergraduates at Bangor University, a university in Wales, UK. EU Law is a compulsory undergraduate course. Most students are in their second year of a three-year programme, while a small minority is in the final year of a shorter two-year undergraduate programme. With 128 enrolled students, the class size was relatively large. At Bangor University, EU Law is a so-called 'short and fat' module taught intensively between late September and the Christmas break with four hours per week. Needless to say, large amounts of complex material must be covered in a short period of time.

The poor reputation of the EU Law module was particularly relevant when designing this teaching intervention. Traditionally, compared to other modules at the same level, EU Law was a module with relatively low average marks. Students said it was very unpopular and complained that they found it difficult to grasp the institutional legal set-up of the European Union. This may not be surprising given that the EU is, indeed, a rather complex (or one could say: messy and bureaucratic) institution and the legal framework is a result of political compromise rather than a domestic legal process to which the students are much more accustomed from their previous studies. The poor reputation of the subject matter was most probably also influenced by the negative reports on the EU in the British mainstream media and official political discourse in the years leading up to the Brexit referendum in 2016. In 2012-2013, I taught on this module for the first time. I was responsible for the teaching of the first half of the semester and I was assigned the role of the module leader. In the UK, co-teaching is common and the module leader assumes the responsibility of the syllabus, exams and all administrative matters related to a course. The Head of School (Dean) taught the second half of the semester. Two postgraduate tutors supported students in smaller-group weekly tutorials, some of them held in Welsh given the importance of bilingualism at Bangor University.

### *2. The Intervention and Its Evaluation*

The intervention consisted of a series of confident-based MCQ quizzes used formatively in the lecture theatre and a final MCQ test used at the summative exam.

*More than a Lucky Guess?*

In order to make the lectures interactive and to allow students to monitor their learning process, I devised three quizzes with each ten (single-choice) categorical items (colloquially »MCQ«) on key concepts taught in the module. I selected questions that students could answer with at least a superficial knowledge of key terms, cases and institutions related to EU law already covered in the previous classes and the textbook. The questions were intended to check basic and superficial learning and would clearly be unsuitable to evaluate the full list of learning outcomes of the module. In subsequent quizzes, I sometimes inserted a slightly modified version of an already used question that still seemed to cause difficulties (such as the distinction between the European Council, the Council of the European Union and the Council of Europe). I used these repetitions to give students another chance to respond correctly as well as to flag and to (re)explain the importance of the issues covered in the question.

Example of a quiz question:

1. What is the most accurate and specific description of the meaning of 'direct effect' of an EU provision?
    - a) The EU provision is of general application and therefore legally binding
    - b) The provision takes precedence over provisions of national law
    - c) Individuals can rely on the provision in the national court
    - d) The provision is binding on national courts from the date of its adoption
    - e) National legal systems have discretion but must achieve a certain result
  2. How sure are you that your answer to question 1 is correct?
    - a) I am convinced
    - b) It is more likely than not that my answer is correct
    - c) I am guessing
- (The correct answer to question 1 is c.)

The three quizzes were used for formative assessment in class during the regular course sessions. Students answered the quizzes individually. This was done in a relaxed non-examination situation. Students were asked to remain silent and to try to answer the questions as good as they could but no other steps were taken to, e.g., prevent them from talking to their neighbours or from consulting their notes. Students filled out the quizzes together with a confidence range. i.e. the student indicated how sure he or she felt about the correctness of the answer. I subsequently read out the correct answer to the

questions orally in class and the students calculated their own scores. While going through the answers, I re-emphasised the key concepts and answered questions from the students. Many if not most students volunteered information (acoustically and in body-language) when their answers were correct and vice versa. This way, the »muddiest points« from the quizzes (i.e. the questions with which a high number of students struggled) often became immediately apparent. I conducted a brief survey of the range of points scored with a show of hands and provided feedback about whether the distribution corresponded to my expectations. Rather than immediately testing the use of confidence-levels in summative exams, I decided to experiment with this idea in formative in-class tests. The summative exam contained a »traditional« MCQ component (a very similar quiz but marked without confidence-levels). While I had assumed that I would collect the formative MCQ quiz sheets from the students, I later decided that the students should keep their own sheets so that they had a chance to go back to the quiz and their own answers at home. Students thus reported the answers on two separate sheets: They circled his or her answers on the printed quiz and kept this sheet. Each student also reported his or her answers on the University's machine-readable TeleForm exam answer sheets which the students handed in. This proved to be a valuable change. There were indeed a few problems with the University-wide MCQ answer sheet when students first encountered this standardised form. A few students had correct answers, but reported them in the wrong place on the form (confusing the horizontal and vertical layout of the form). These mistakes allowed me to clarify the issue in time for the second quiz and was probably crucial to avoid potential confusion at the exam (with a potential risk of complaints given the unfamiliar nature of the assessment component). There were no problems with the answer sheet in the summative exam. This is how the formative in-class quiz scores were calculated:

<b>Correct</b>	<b>1</b>
convinced.	+3
... more likely than not that my answer is correct	+1
guessing.	+0

<b>Wrong</b>	<b>0</b>
convinced.	-3
... more likely than not that my answer is correct	-1
guessing.	-0

Examples: Student A answered a question correctly but admits that he was guessing. Correct answer = 1 point; Guessing = 0 points = total of 1 points. Student B answered a question wrongly and was convinced that the answer

### *More than a Lucky Guess?*

would be correct: Wrong answer = 0 points; Convinced = -3points = a total of -3 points for this question.

The evaluation consisted of several elements. First, I evaluated the collected MCQ sheets to monitor learning progress. If students struggled with a concept, I explained it again immediately after the quiz and in the subsequent week (recap section at the beginning of class) and considered including a similar question in the next MCQ. Second, at the end of the module, I collected feedback from the students in a survey, evaluated exam results and the so-called Arque report (Assessment Reports on the Quality of University Examinations, a detailed statistical report).

## *II. Summary of Relevant Data and Feedback*

The following empirical data is available:

### *1. Scores of MCQ Quizzes*

The three quizzes were filled out in the second and fourth week of the semester as well as during the revision lecture at the end of the semester. In the second week of term, out of 128 enrolled students, 94 submitted their answer sheet. In the fourth week, 75 answer sheets were collected.

Unfortunately, the timing of the revision lecture was far from ideal in the late afternoon before Christmas. Attendance was very poor and only 36 students were present and handed in their answer sheets.

The in-class quizzes showed normal but widely spread Gaussian distribution, with few students scoring very few or very many points and an accumulation in the centre. At the summative exam, many students scored well (8-10 out of 10 points) and the distribution was thus left-skewed (or negatively-skewed) with a long-left tail with few students scoring poorly and a high number of students scoring very well.

### *2. Student Feedback*

Student feedback consisted of a small survey with closed and some open evaluation questions. As mentioned above, only a small group of the cohort attended the revision class in which the feedback data was collected. Written comments to the open evaluation questions in the revision lecture were extremely scarce: out of the 36 students attending, only five students provided

scant comments and it is difficult to analyse these few comments in any systematic way. It can be assumed that only »lovers and haters« took the time to fill out the comment box before leaving for Christmas. Those who commented were either enthusiastic or, in contrast, two students thought the MCQ took up too much time. 30 out of 33 students found the exercise was »useful« or »rather useful«, 32 found the exercises a good way to receive individual feedback to identify the aspects of EU law for which I need to do more readings and revision and those for which they can already feel confident about. 33 out of 35 students answered that they went back to the course material after the MCQ exercises and a vast majority (25/33) answered that they enjoyed the quizzes.

### *3. Statistical Data from the Summative Exam*

At Bangor Law School, each lecturer receives a statistical report on the assessment of each module and is then asked to comment on any statistically unexpected results at the annual exam board or in a conversation with the external examiners and/or the Head of School. Before a grade is final, the external examiners (one or several teachers from another university) satisfy themselves that assessments are conducted in accordance with the university's quality standards. These so-called Arque reports contain an overview with graphs and comments for each module. The picture below contains the Arque report for the EU Law class in which the teaching intervention took place.<sup>5</sup> It is only of partial use to evaluate the teaching intervention given that it contains data on the summative assessment as a whole rather than merely on the MCQ part. Nevertheless, it is included here as the inclusion of the ten multiple-choice questions did seem to influence the outcome of the final assessment to some extent.

Selected extracts from the statistical Report for Module SXL2110 EU Law, taken from the Bangor University Examination Reports, Law, 18 June 2013:

<sup>5</sup> The final exam was taken by a total of 115 students. The class size of 103 is explained by the fact that this report only relates to those students who were enrolled in the ordinary three-year BLaw programme and excludes the students of the much smaller two-year programme.



*More than a Lucky Guess?*

Class size	103
Marks missing	0
Zeroes	1
Mean	53.9
SD	11.7
Pass-rate (%)	89.2

Figure 1: basic statistical data for the EU law marks (zero marks are excluded in the graphs below)

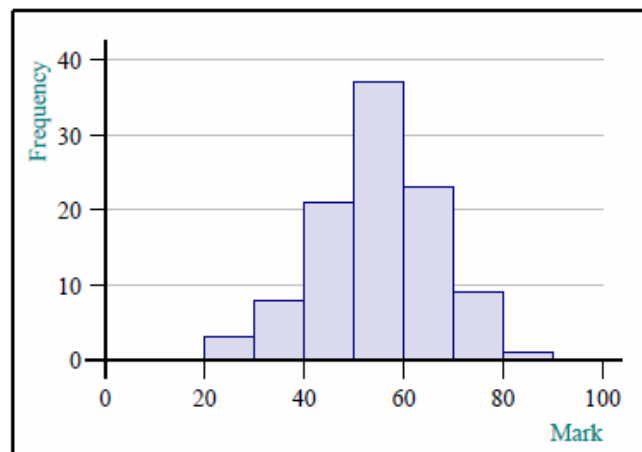


Figure 2: distribution of marks (threshold to pass = 40)

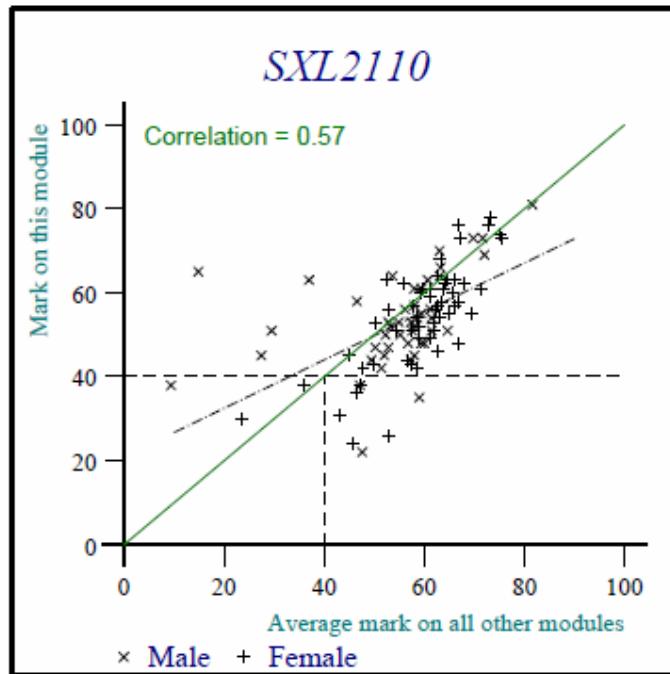


Figure 3: correlation indicating the degree of association between each student's mark in the EU law module and his/her average mark on all other modules.

Interestingly, the Arque report compares students' individual mark with the same individual's average scores on other modules during the same year. We would expect that students scoring well in other modules would also perform well in EU Law and poor students would score comparatively poorly (resulting in mean differences of approximately zero). The Arque report revealed that this was not entirely the case: «The correlation of the scores on this module with average scores on other modules is 0.57. The regression line of score on average score has a slope of 0.6. This is statistically significantly smaller than a slope of 1 at the 5% level. (The p-value is <0.01)» This raises the question whether the assessment discriminated adequately across the ability range. This issue is discussed below in the discussion section.

### *III. Discussion*

#### *1. The Rationale for the Intervention*

Biggs famously suggested that students' approaches to learning and studying can be modified by the teaching and learning context.<sup>6</sup> How could confidence-based MCQ quizzes influence the students' learning approaches in a way that would enhance their understanding of the material?

##### *a. Active Participation, Feedback, Motivation and an Opportunity for Self-Assessment*

The learning of introductory factual knowledge can be encouraged by quizzes,<sup>7</sup> and without acquiring basic factual knowledge, none of the other learning outcomes of the module can be achieved. Students generally seem to like quizzes and the atmosphere in the lecture theatre tends to relax significantly with a quiz. Although the primary goal of higher education is not to induce entertainment or relaxation, emotions play a proven role in enhancing or hindering student motivation.<sup>8</sup> Positive emotions and the feeling of a joint enjoyable experience in the classroom can reduce fear and contribute to a positive relationship between students and lecturer. In any event, novelty and variation increase students' (situational) interest.<sup>9</sup> Ann Morton emphasises the well-known challenge that breaking the flow or changes activity in large lecture rooms is crucial to generate active participation and overcome the problem of the very limited human capacity to concentrate on a lecture »where the lecturer is monotone, the PowerPoint presentation is a predictable list of bullet points and at no point do [the

<sup>6</sup> Biggs, *Teaching for Quality Learning at University* (fn. 1), p. 39 (39-45).

<sup>7</sup> Klappa, *Promoting Active Learning through »Pub Quizzes« — a Case Study at the University of Kent*, in: *Bioscience Education* 2009 14/1, p. 1.

<sup>8</sup> Zumbach/Astleitner, *Effektives Lehren an Der Hochschule: Ein Handbuch Zur Hochschuldidaktik*, Stuttgart 2016, p. 34. On the importance of fear as a significant hindrance of learning, see in particular Pekrun/Götz, *Emotionsregulation: Vom Umgang Mit Prüfungsangst*, in: *Handbuch Lernstrategien* 2006, p. 248.

<sup>9</sup> Jones, *Motivating Students to Engage in Learning: The MUSIC Model of Academic Motivation*, in: *International Journal of Teaching and Learning in Higher Education* 2009 21/2, p. 272 (278).

students] feel part of the lecture«. <sup>10</sup> She suggests that MCQ »mini-tests« to check student progress are one out of many possible ways to increase student participation and variation. <sup>11</sup> Surely, there is nothing controversial about this proposition. But it doesn't tell us much yet about how such MCQ quizzes should be used with the best prospects for learning.

In particular, David Nicol conducted a meta-study on best-practices related to the use of multiple-choice tests. He concludes that MCQs can be used to support the development of learner autonomy, but that test construction is not sufficient to do so. Rather, *the context* in which these tests are used is highly significant. <sup>12</sup> Importantly, he emphasises how MCQ quizzes can be used to provide feedback which complies with the 'seven principles of good feedback' while still manageable with large groups. <sup>13</sup> Amongst other criteria, feedback must be timely and individual, <sup>14</sup> but providing such feedback can be challenging when the class size is relatively large. Crucially, MCQ allow for instantaneous feedback without an onerous marking load. <sup>15</sup> Furthermore, in-class MCQ quizzes can link formative tasks to summative assessments and thus help to clarify goals and standards and can help shaping teaching by identifying areas of difficulty. Research shows that giving repeated opportunities to take MCQ tests is motivating and that this motivation can further be enhanced »when this formative procedure is linked to later summative tests of a similar format«. <sup>16</sup> Importantly, the intervention moreover allows students to develop capacity for self-evaluation given that ungraded tasks leading to summative assignments »help students to self-assess their learning needs« <sup>17</sup> and given that the MCQ were marked with confidence-levels whose rationale is further explained next.

<sup>10</sup> Morton, Lecturing to Large Groups, in: Fry et al. (ed.), *A Handbook for Teaching & Learning in Higher Education*, London 2003, p. 58 (64).

<sup>11</sup> Morton, Lecturing to Large Groups (fn. 10), p. 58 (64).

<sup>12</sup> Nicol, E-Assessment by Design: Using Multiple-Choice Tests to Good Effect, in: *Journal of Further and Higher Education* 2007 31/1, p. 53 (53-64).

<sup>13</sup> Nicol, Using Multiple-Choice Tests to Good Effect (fn. 12), p. 53. See also Nicol/Macfarlane-Dick, Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice, in: *Studies in Higher Education* 2006 31/2, p. 199 (199-218).

<sup>14</sup> Race, *The Lecturer's Toolkit: A Practical Guide to Assessment, Learning and Teaching*, London; New York 2007, p. 79.

<sup>15</sup> Harley, et al., High-Level Multiple Choice Questions in Advanced Psychology Modules, in: *Psychology Learning & Teaching* 2009 8/2, p. 30 (8, 30).

<sup>16</sup> Race, *The Lecturer's Toolkit* (fn. 14), p. 56.

<sup>17</sup> Liu, et al., Evaluating Oneself, in: Carless (ed.), *How Assessment Supports Learning Learning-Oriented Assessment in Action*, Hong Kong 2006, p. 125 (125, 137).

*More than a Lucky Guess?*

*b. The Specific Rationale of Using Confidence-Levels*

Rather than relying on traditional MCQ, a more innovative approach stemming from the medical sciences also asks students to rate their own answers according to how confident they feel about them. The idea is that the indication of a confidence level allows the student to obtain more meaningful feedback: Students identify points of weakness and develop critical self-awareness of which parts of their knowledge are sound and which are not – which in turn assists further learning.<sup>18</sup>

The main reference for the scholarly evidence on the use of confidence-levels in multiple-choice tests is Tony Gardner-Medwin, now emeritus professor of physiology. Gardner-Medwin analysed whether confidence-based marking could be used »towards deeper learning and better exams«. <sup>19</sup> He used confidence-based marking at the University College London in undergraduate medical exams and enthusiastically reported the benefits of this approach. In particular, he stresses that confidence-based marking places a premium on careful thinking: »It ... encourages deeper understanding and learning. It is popular with students, and helps them develop valuable skills. In exams, it produces higher quality data for their assessment.«<sup>20</sup>

Adams and Ewen criticise that there is a »guesswork factor« involved in traditional MCQ assessments.<sup>21</sup> Yet, in the workplace (or in life in general), the performance of individuals is strikingly different between those who guess correctly and those who really know that something is correct.<sup>22</sup> Confidence-based marking distinguishes lucky guesses from true knowledge and contributes to more reliably discriminating across the ability range.

Against the background of the poor reputation of the EU Law module, the confidence-based scoring allowed students to see that some of their answers were correct – sometimes more than they had expected. In this process, they

<sup>18</sup> Gardner-Medwin, Confidence-Based Marking – Towards Deeper Learning and Better Exams, in: Bryan/Clegg (ed.), Innovative Assessment in Higher Education, London; New York 2006, p. 141.

<sup>19</sup> Gardner-Medwin, Confidence-Based Marking (fn. 18), p. 141 (141).

<sup>20</sup> Gardner-Medwin, Confidence-Based Marking (fn. 18), p. 141 (149).

<sup>21</sup> Adams/Ewen, The Importance of Confidence in Improving Educational Outcomes, Presented at the 25th Annual Conference on Distance Teaching & Learning, Madison 2009, <[http://www.uwex.edu/disted/conference/Resource\\_library/proceedings/09\\_20559.pdf](http://www.uwex.edu/disted/conference/Resource_library/proceedings/09_20559.pdf)> (last visited: 2 March 2017).

<sup>22</sup> In this regard, see also Bruno, Using Testing to Provide Feedback to Support Instruction: A Reexamination of the Role of Assessment in Educational Organizations, in: Item Banking: Interactive Testing and Self-Assessment, Berlin 1993, p. 190.

were cued to believe that competence in EU Law can be improved and the subject matter can be studied just like their other law subjects. I felt this was particularly relevant in a context in which a subject matter has the reputation to be hopelessly unintelligible and students begin their classes with a »failure orientation« and low expectations about their own learning.<sup>23</sup>

### *c. Ensuring Comprehensive Revision Behaviour*

A secondary consideration that motivated me to experiment with MCQs was my intention to disincentive selective learning. Students in Bangor Law UG modules must be given a wide range of choice as to the selection of essay and problem questions in exams. Many students seemed to strategically prioritise only very few selected areas for revision. Any reasonably smart student could look up the past exams on the website of the library, identify a few areas covered in all previous exams and then decide to adopt a »surface approach« to retain enough information for these areas so as to pass the exam without engaging with any other part of the material. This impression was based on my undergraduate teaching on other modules and conversations with the previous module leader for EU Law. Such a rational but very selective revision behaviour leads to a very perfunctory approach to learning.<sup>24</sup> Including a compulsory MCQ component in the exam and ensuring that the MCQ questions cover the key concepts from all topics of the module would provide incentives to revise a wider range of the material. The MCQ component enhances the teaching-assessment alignment and encourages students to acquire at least a basic understanding of all key areas of the module rather than three or four areas they speculate would come up in the exam studied in isolation from each other.

<sup>23</sup> When students believe that competence cannot reasonably be improved, their learning is very seriously hindered. *Steele-Johnson/Beauregard*, Goal Orientation and Task Demand Effects on Motivation, Affect, and Performance, in: *Journal of Applied Psychology* 2000 85/5, p. 724 (725). See also *Astleitner*, *Aufgaben-Sets und Lernen: Instruktionspsychologische Grundlagen und Anwendungen*, Frankfurt am Main 2006, p. 75 (writing that relatively simple tasks are useful to overcome low self-confidence).

<sup>24</sup> On the influence of assessment methods and the learning context on students' approaches to study, see also *Fry, et al.* (ed.), *A Handbook for Teaching & Learning in Higher Education*, London 2003, p. 10-12.

## *More than a Lucky Guess?*

### *2. Discussion of Statistical Results at Summative Exams*

As mentioned above in the summary of relevant data, the examinations report showed a lower slope for the regression line of score on average score than what would statistically be expected. A slope = 1 of this curve means that students who are generally strong across modules are equally strong in EU law and vice versa for weak students. I would speculate that the slope < 1 for this course was due to a combination of three factors, one of them relating to the use of a multiple-choice component at the summative exam:

First, there were a number of outliers in the upper left quadrant of the graph. If these are not taken into account, the slope would be closer to 1. Second, a relatively high number of average students scored just slightly less in EU law than they do in the other modules. They account for the many dots just below the slope = 1 line in the centre of the graph. The effect of this phenomenon is that the curve shifts clockwise. This corresponds to the experience with EU law over several years: grades are historically slightly lower. Third, and importantly, the fact that some otherwise relatively weak students scored above their individual averages might be explained by the use of the new assessment method by MCQ. Analytically weak but diligent (or lucky) students could improve their exam marks with good revision and retention of the key concepts required to score well on the MCQ questions. This might explain why some of the weaker students scored better than they do in their other modules – which again would have the effect of shifting the regression line clockwise, i.e. reducing the slope of the regression line.

Overall, I did not think that the lower slope was a problem given that the standard deviation was still comparable or slightly higher than the standard deviation in other large modules and the grade distribution as a whole follows a normal Gaussian distribution. However, it would have been necessary and interesting to observe the behavior of this regression line if confidence-levels would have been used in the subsequent year. This data is unfortunately not available as I did not teach the module in the next year.<sup>25</sup>

### *3. Concerns about Using Confidence-Based MCQ*

There is a major concern about using confidence-based MCQ and this concern deserves to be taken seriously.

There is sometimes a perception that confidence-based marking would unfairly disadvantage risk-adverse personalities (supposedly more common

<sup>25</sup> I subsequently returned to Switzerland.

among female students). Gardner-Medwin's research shows that there might indeed be such an effect when students first encounter confidence-based marking. However, he could find no evidence in his own research. Nevertheless, he points out that »if, quite plausibly, individuals or groups do have initial tendencies to be under- or over-confident then this is an objective problem that [the students] should be aware of. Confidence-based marking offers suitable feedback and training. (...) in decision-rich occupations such as medicine, miscalibration of reliability is a serious handicap.«<sup>26</sup> I believe that the same is true in law: a lawyer who is overconfident might provide misinformation to a client. On the other hand, a lawyer who is under-confident engages in unnecessary additional research, might not succeed at oral advocacy despite good arguments and/or might not inspire confidence and trust from the client and his or her peers. Over- and under-confidence thus come with serious professional disadvantages. Hence, Gardner-Medwin acknowledges that the bias may exist, but thinks that it is preferable to make student aware of their own tendency to be under- or over-confident so that they can self-assess their confidence more accurately or at least know the risks they are taking. I felt that this potential bias, in any event, wasn't a problem if I only experiment with confidence-levels in a formative setting rather than in the summative exam. However, I do think that it is justified to accept this potential problem even in a summative exam setting – for the reasons given by Gardner-Medwin above if, but only if, the problem is explicitly mentioned in class and students are given appropriate opportunities to familiarise themselves with the use of confidence-levels. Using confidence-based MCQ quizzes repeatedly has the additional benefit of task predictability and may thus reduce fear.<sup>27</sup>

#### *IV. Reflections on the Implications for Future Teaching Practice*

Overall, the use of confidence-based MCQ quizzes seemed to render significant benefits. In particular, the MCQ quizzes were interactive, beneficial for the atmosphere in the lecture theatre and effective in increasing the attention to the solutions of the quiz as well as useful to inform my recap sequences at the beginning of new lectures. I felt it was relatively time-consuming to design the quizzes, but not to implement the in-class activity. Although the empirical data collected has weaknesses, it is reasonable to believe that the quizzes reduced fear and increased students' confidence and motivation towards the new and dreaded subject matter.

<sup>26</sup> Gardner-Medwin, *Confidence-Based Marking* (fn. 18), p. 141 (146-147).

<sup>27</sup> Astleitner, *Aufgaben-Sets und Lernen* (fn. 23), p. 32.



*More than a Lucky Guess?*

Some lessons were learned for improvement. Subsequent attempts should use better evaluation methods and sequencing. To generate more and more meaningful feedback specifically on the use of confidence-based MCQs, an intermediary evaluation should be done early on and then another one after the summative exam when the students have obtained their exam results. Apart from the fact that the collection of student feedback did not work according to plan, it would also be interesting to collect more detailed data at the baseline and on emotional effects of confidence-level MCQ, using more elaborate methods. At the Salzburg conference, the promising idea of using Tandems was presented in order to overcome methodological challenges. An educational scientist or a psychologist could pair up with a law lecturer to design and implement better evaluation methods.

Furthermore, it would be ideal to study the use of confidence-level MCQ over the course of a longer period, i.e. during several years. Given technological advances, it would also make sense to take advantage of technological developments in the area of e-assessment, such as dedicated exam software (such as EvaExam-EvaSys) or Electronic Voting Systems.<sup>28</sup>

I later also realised that I did not comply with Gardner-Medwin's advice on the scoring scheme. Particularly if used in a summative exam situation, it would be crucial to use neutral descriptors of confidence (low, mid, high) rather than descriptive terms (such as certain, very sure, unsure, guess, etc.). Gardner-Medwin rightfully warns that »descriptions mean different things to different people and in different contexts. A transparent mark scheme must be defined by rewards, penalties and explicit risks, not by subjective norms.«<sup>29</sup> In addition, Gardner-Medwin mathematically tested the effect of various marking schemes to ensure that it doesn't pay to misrepresent one's confidence. The ideal mark scheme for all question types with more than two possible answers is the following:

<i>Confidence level</i>	<i>C=1 (low)</i>	<i>C=2 (mid)</i>	<i>C=3 (high)</i>	<i>No reply</i>
Mark if correct	1	2	3	0
Penalty if wrong	0	-2	-6	0

*Table 1: The »London Agreed Protocol for Teaching« confidence-based mark scheme, from Gardner-Medwin.<sup>30</sup>*

<sup>28</sup> On the use of in-class electronic voting systems, see the insights by *Draper*, *Catalytic Assessment: Understanding How MCQs and EVS Can Foster Deep Learning*, in: *British Journal of Educational Technology* 2009 40/2, p. 285.

<sup>29</sup> *Gardner-Medwin*, *Confidence-Based Marking* (fn. 18), p. 141 (142).

<sup>30</sup> *Gardner-Medwin*, *Confidence-Based Marking* (fn. 18), p. 141 (142).

Taking the above-mentioned lessons into account, I would encourage others to experiment with confidence-based MCQ and I am looking forward to learning from their experiments with this approach.