Assessment in Biology: Trends, Problems and Solutions

Fiona L. Bird

Corresponding author: f.bird@latrobe.edu.au
Department of Zoology, La Trobe University, Bundoora VIC 3086, Australia

Keywords: formative assessment, summative assessment, feedback, critical thinking, assessment practices

International Journal of Innovation in Science and Mathematics Education *22*(2), 85-99, 2014.

Abstract

Our challenge as biology educators is to design assessment practices which support high quality student learning and result in knowledgeable, critically thinking students who can self-regulate their own learning. A review of the last decade of literature was conducted with the aim of quantifying the type and content of articles written about assessment in biology, identifying the problems currently being investigated by biology educators and discussing some of the solutions being offered and evaluated. The quantitative review found that exams/tests were the most frequently discussed assessment practice and the theme of guidance and feedback provision dominated the research. The major problems addressed in the literature were designing exam questions to better assess higher level cognitive skills, creating opportunities for formative assessment of content, developing critical thinking skills and developing the capacity of students to self-regulate their learning. The review revealed that biology educators have been rigorously reflecting on and evaluating assessment practices, and have implemented curriculum reforms to address these problems. The challenge for the future will be to encourage broader adoption of these assessment practices by academics so that biology students everywhere will graduate well-equipped for the workforce.

Introduction

In their document, *Assessment 2020: Seven propositions for assessment reform in higher education*, Boud and Associates (2010) state that *"Universities face substantial change in a rapidly evolving global context. The challenges of meeting new expectations about academic standards in the new decade and beyond mean that assessment will need to be rethought and renewed."* In addition to calls for reform in assessment across the university sector, the teaching of science has been under the microscope with calls for repositioning content knowledge across the science disciplines. The rapid growth in knowledge in the discipline of biology has led to questions about the importance of facts and a shift towards an emphasis on equipping students with the skills and knowledge to find, analyse, synthesise and apply information to new situations and challenging problems goals (e.g., Brewer & Smith, 2011). At the same time, a growing body of work around standards of achievement for science graduates (e.g., Jones, Yates & Kelder, 2011 (Australia), European Tuning Project, UK Quality Assurance Agency) has resulted in a clear articulation of the learning outcomes science graduates are expected to achieve. These include skills and competencies around understanding science, inquiry and problem-solving, communication, and personal and professional responsibility, alongside a broad and deep disciplinary knowledge. In combination, these major changes in thinking about science and biology education are anticipated to spark curriculum renewal as educators consider and implement new initiatives.

Assessment will be a key element in this curriculum renewal, and any changes should be made with reference to the substantial body of knowledge about effective assessment practices documented in the higher education literature. The purpose of this review is to document the current trends in research about assessment in biology, rather than repeat the currently accepted principles of assessment design (see Boud & Associates, 2010 for an overview). This review is not intended to provide a list of all assessment tasks related to biology. Comprehensive databases of resources, such as the website already fulfil that role.

Specifically, this review aims to quantify the type and content of articles written about assessment in biology, identify the problems currently being worked on by biology educators and discuss some of the solutions being offered, trialled and evaluated.

The literature review

The ISI Web of Science database was searched using combinations of the search terms biology, education, assessment, critical thinking, rubric, laboratory skill, self-assessment and peer assessment. The search was limited to peer-reviewed journal articles published in the past decade (2002-current) with a focus on undergraduate assessment practices in the discipline of biology (not including health sciences, medicine, dentistry or pharmacy). The database search was complemented by a direct search of the last decade of back issues of the key journal *Assessment and Evaluation in Higher Education*. The combined database and direct search revealed 59 journal articles.

The content of each journal article was categorised in three ways: genre, assessment theme and nature of the task. Hounsell, McCune, Hounsell and Litjens (2008) completed a census of innovative assessment being reported in the United Kingdom (UK) across the time period 1997-2007 and the genres and categories identified in the UK study have been used as a basis for the quantitative analysis of journal articles in the current review. Any comparison between the two datasets was limited as the UK review only included articles about 'innovative' assessment practices whereas the current review included articles about any form of assessment in biology.

Genre

In the last decade, 35 articles (60%) written about assessment in biology were *empirical studies*, 20 were *accounts of practice* (34%) and four were *commentary/opinion* and *theory/conceptualisation* pieces (3%) (Table 1). This contrasts with the findings of Hounsell et al. (2008) who found that *accounts of practice* dominated the UK literature in the decade preceding 2007 (58% of the 17 papers documented in the review). The relatively greater proportion of empirical studies in the current review may be in part due to regional differences in approaches to higher education research (this review included the international literature compared to the UK focus in Hounsell et al., 2008), but also may be in response to calls for more rigorous evaluation and reporting of curriculum reforms (e.g., Sundberg, 2002).

Nature of the assessment task

The most frequently discussed assessment task was the test or exam (including constructed response (short answer or essay) and multiple choice questions (26 papers)) (Table 2). Non-exam written assessment tasks included laboratory, research or project reports (13 papers) and alternatives such as essays, question sheets, article critiques (11 papers). Oral or poster presentations were also frequently discussed (10 papers). Assessment practices that received the least attention in the literature included laboratory and scientific skills (4 papers), case study assignments (2 papers) and group work (2 papers). It is clear from this analysis that exams and written tasks still dominate assessment practices in biology, although oral/poster presentation of research was also discussed frequently. Assessment of group or teamwork was less prevalent, with it being mentioned only twice in the 59 articles considered by this review. This trend is of particular concern because teamwork is a personal and professional skill highlighted as a threshold learning outcome in science (Jones et al., 2011). Students will not become accomplished team-players if they are not given formative and summative assessment opportunities in that skill throughout their undergraduate studies.

Table 1. Categorisation of journal articles by genre (after Hounsell et al., 2008)

Genre	N
Empirical Study A publication that aims to report the findings of a research enquiry, investigation, experiment or survey of assessment practices, processes or policies	35
Account of practice A publication that aims to describe and reflect on an instance of a change or development in day-to-day professional practice in assessment, and is usually self-reported and self-evaluated by one or more subject practitioners.	20
Commentary/Opinion Piece (including a contribution to debate) A publication that aims to argue for a reappraisal of, or a significant change to, one or more aspects of contemporary assessment practices, processes or policies.	2
Theory/Conceptualisation A publication that seeks to contribute to theoretical advance, whether by proposing a new conceptualisation of assessment processes, practices or policies, or by modifying or refuting an existing one.	2
Total	**59**

Table 2. Categorisation of journal articles by nature of the assessment task. Where a publication falls into more than one category it has been counted under each of the relevant categories.

Nature of the task	N
Test or exam question (constructed response or multiple-choice)	26
Laboratory or research report	13
Other writing assignment	11
Oral or poster presentation	10
Scientific and laboratory skills	4
Case study assignment	2
Group work and collaboration	2

Themes

The theme of *guidance and feedback in assessment* was present in 31 articles (53%) and s*tudent involvement in assessment* was discussed in 13 articles (22%) (Table 3). Similar numbers of articles focussed on *modes of assessment* (eight articles), *assessment criteria* (eight articles), *perceptions and experiences of staff and students* (seven articles), *enhancing assessment practice* (six articles) and *use of new technology in assessment* (five articles). Seven articles included in this review mentioned assessment as part of a new curriculum, but evaluated the curriculum as a whole rather than focussing on the assessment. These articles were categorised as 'no theme'. A detailed list of topics and references listed under each theme can be found in Appendix I.

Table 3. Categorisation of journal articles by assessment theme. Where a publication falls into more than one category it has been counted under each of the relevant categories.

Theme		N
1.	What guidance and feedback is given, by what means and when	31
2.	Student involvement in assessment	13
3.	Modes of assessment and balance between them	8
4.	Criteria used and their relative weighting	8
5.	Perceptions and experiences of staff	7
6.	Enhancing assessment practices	6
7.	Use of new technology in assessment	5
8.	No theme	7

In summary, the dominant genre was the empirical study. Test and exam questions were the most frequently discussed assessment tasks with most papers focussed on guidance and feedback. The next section of this review will draw out in more detail some of the problems with assessment identified by biology educators and strategies to improve assessment practice. The problems discussed in the next section were identified in the collective literature of this review, however, the solutions are drawn primarily from *empirical studies* and the *commentary/opinion* and *theory/conceptualisation* pieces rather than *accounts of practice*. *Accounts of practice* are valuable for sharing new ideas about teaching and learning activities, however, *empirical studies* add to the collective knowledge about assessment as a whole and the practices they evaluate can be readily applied to other settings.

Problems and solutions identified in the literature

Improving the cognitive level of exam and test questions

Tests and exam questions were the most frequently discussed assessment items in the assessment literature in biology in the last decade. Multiple choice questions, in particular, were identified as troublesome in several articles, even though this question type is still the most frequently used assessment in biology (Goubeaud, 2010). A recent survey of 1535 academics from a range of public and private American higher education institutions confirmed that 73% used multiple choice exams in all or some of their classes (Goubeaud, 2010). Another study of 77 introductory biology subjects from across the United States of America found that the types of questions asked in high-stakes quizzes and exams overwhelmingly targeted lower cognitive levels and were often at odds with the levels of difficulty articulated by academics in course goals (Momsen, Long, Wyse & Ebert-May, 2010). A curriculum assessed mainly by multiple choice questions testing recall is likely to encourage adoption of surface learning approaches, such as memorising facts and processes, rather than critical engagement with the content (Ramsden, 2003). If our goal is to train critically thinking scientists, then we need to design assessment items which facilitate deep learning of the material. In the current higher education environment of increasing class sizes and decreasing teaching budgets, multiple choice exams remain a cost-effective assessment method. It is therefore increasingly important that if multiple choice items are used in exams and quizzes that they effectively test critical thinking as well as content.

There are tools available to assist biology educators to design questions which assess critical thinking and higher cognitive level skills. The Blooming Biology Tool (BBT) uses Bloom's taxonomy to analyse the cognitive level that a question addresses (knowledge, comprehension, application, analysis, synthesis or evaluation) and to design items so that they intentionally assess the desired cognitive level (Crowe, Dirks & Wenderoth, 2008). The study found that the BBT assisted staff to write questions targeting higher cognitive level skills and thereby assisted their students to better develop these skills (Crowe et al., 2008). Bissel and Lemons (2006) also used Bloom's taxonomy to inform design of constructed response assessment items in an introductory biology course, and developed rubrics for marking the items. The rubrics allowed for independent assessment of both the content and the critical thinking skills required to answer each question correctly, thereby highlighting to students the need for higher level cognitive skills to correctly answer the questions (Bissel & Lemons, 2006). Richmond, Merritt, Urban-Lurain and Parke (2010) developed a tool which focussed on the role of models in cellular biology and the skills associated with understanding, constructing and applying biological models. Like the BBT, this tool was used to analyse existing assessment items about biological models, in terms of the cognitive level assessed, and revise or design new items to better align with the curriculum goals (Richmond et al., 2010). Stanger-Hall (2012) encouraged critical thinking by students about the content of an introductory biology subject by providing practice constructed response questions (such as short answer or essay) and incorporating these types of questions into what was a traditional multiple choice exam. The study found that when students knew that they would be assessed with constructed response questions, they adopted more cognitively active study behaviours which resulted in a significantly better performance in the final exam (Stanger-Hall, 2012).

Creating opportunities for formative assessment of content

The two themes of *guidance and feedback* and *involving students in assessment* dominated the literature examined in this review. These two themes are particularly pertinent to the second problem of providing students with feedback on their learning of content prior to the final summative assessment. In his review of innovations in teaching undergraduate biology, Wood (2009) identified in-class active learning tasks as an effective way to provide immediate feedback on learning to students and staff and reinforce student engagement with content throughout the semester. Several recent studies have found clear evidence that student engagement and learning improve after the introduction of these new formative strategies. Classroom response systems (or clickers) have been used to conduct quizzes in class to give students instant feedback on their learning (Preszler, Dawe, Shuster & Shuster, 2007; Freeman, Haak & Wenderoth, 2011; Ueckert, Adams & Lock, 2011). Preszler et al. (2007) investigated student performance on exam questions assessing material from lectures with small, medium and large numbers of in-class questions and found that the increased use of clickers in class positively influenced performance of students across six biology subjects (Preszler et al., 2007). Freeman et al. (2011) compared student performance in a highly structured course (reading quizzes, and/or in-class active learning activities such as clickers and weekly practice exams) with a low-structure course (lectures and a few high-stakes assessment) and found that failure rates were substantially smaller in a highly structured course. Haigh (2007) found that quiz results were significantly correlated with exams and oral presentations which required immediate individual knowledge, but not with assessment tasks involving group work or problem solving. This indicates that while formative quizzes may assess facts and concepts, they do not assess student performance in achievement of higher cognitive level outcomes. Improved learning outcomes could be gained if in-class quiz questions were designed to also assess the application of this knowledge to new scenarios.

Two studies described formative assessment activities where students actively engage with the content by constructing multiple choice questions (Crowe et al., 2008; Bottomley & Denny, 2011). Crowe et al. (2008) found that formal training of students to use the Blooming Biology Tool for evaluating and categorising questions, followed by repeated opportunities to practice writing and reviewing questions at the different levels of Bloom's Taxonomy, improved study skills and students' capacity to think critically. A study by Bottomley and Denny (2011) showed that second year biomedical science students wrote high quality multiple choice questions, indicating that they were engaging deeply with the material to design questions that were novel, correct and devoid of misconceptions. Students are known to have problems recognising the difficulty of a test item and therefore the cognitive level at which the subject matter assessed in that item, needs to be learned (Lingard, Minasian-Batmanian, Velaa, Cathers & Gonzalez, 2009). Ultimately, such lack of expertise may compromise exam preparation and performance. Formative exercises, where students write and critique test questions, may improve their critical thinking skills and their capacity to recognise the level of difficulty of assessment items.

Constructed response or short answer test/exam questions can be useful in formative and summative assessment. Scoles, Huxham and McArthur (2013) provided annotated exemplars of past exam questions to students and found that the students who accessed the exemplars online performed better in the final exam. Huxham (2007) set coursework questions during semester and provided feedback to students in two ways: personal feedback on their answers as well as model answers to the questions. He found that even though students preferred the personal feedback they received, they performed better using model answers. These studies clearly show that model answers and exemplars provide formative feedback and assist student learning.

Forms of assessment which develop critical thinking skills in students

This review has discussed strategies for developing critical thinking skills in students using well-designed multiple choice and constructed response questions. Several other studies, however, have explored the development of critical thinking about biology using other forms of assessment. Well-developed critical thinking skills are at the foundation of the practice of science and essential for biology students (Brewer & Smith, 2011). There are many examples of authentic inquiry-rich, biology research experiences for undergraduate students described in the literature and most of these papers focus on the benefits of these curricula for the development of critical thinking skills (e.g., Quitadamo, Faiola, Johnson & Kurtz, 2008; Dymond, Scheifele, Richardson, Lee, Chandrasegaran, Bader & Boeke, 2009; Gliddon & Rosengren, 2012). Quitadamo et al. (2008) is one of the few studies which compared an inquiry learning model approach to a traditional lecture/laboratory and specifically quantified the critical thinking gained by students in the same biology subject.

Quitadamo et al. (2008) measured significantly greater learning gains as a result of the types of learning activities and assessments that students experienced using the inquiry model. In the inquiry model students conducted a research project, including initial and final research proposals (with feedback provided on the initial submission), a personal research journal, an oral presentation and a defence of their research at a poster presentation session (Quitadamo et al., 2008). These assessment tasks are authentic artefacts of research that require written descriptions, interpretations and reflections on the research.

It has been suggested that students develop critical thinking skills in part by writing about their research. When a student writes about their research, he or she has to make ideas explicit and choose words and phrases which convey the meaning they intend. Libarkin and Ording (2012) showed that critical thinking skills can also be developed through a set of successive short written assessment tasks. Their study found that when students complete written tasks, they improved their ability to identify relevant background material, incorporate data into building an argument and draw conclusions about the meaning of that data. These are important in the process of scientific thinking developed by the writing process (Libarkin & Ording, 2012). Quitadamo and Kurtz (2007) confirmed that critical thinking skills could also be gained by collaborative writing exercises. Their study found that students who wrote about laboratory exercises had improved skills in analysis and inference compared to students who completed a quiz about the laboratory exercises. The authors argue that the action of writing and the thought processes associated with constructing and supporting an argument or explanation was the learning activity which resulted in the measured gains in critical thinking (Quitadamo & Kurtz, 2007).

Rubrics, which assess scientific reasoning skills in undergraduate writing, have been found to assist academics to explicitly teach and assess critical thinking in writing tasks in biology. Timmerman, Strickland, Johnson and Payne (2011) developed and tested the *Rubric for Science Writing* on a sample of undergraduate biology reports and found that the rubric reliably assessed student achievement of critical thinking and scientific inquiry. The rubric clearly articulates all the criteria for critical thinking skills which students need to develop in an undergraduate biology course under the following headings: Introduction (context, accuracy & relevance), Hypotheses (testable & consider alternatives, scientific merit), Methods (controls & replication, experimental design), Results (data selection, data presentation, statistical analysis), Discussion (Conclusions based on data selected, alternative explanations, limitations of design, significance of research), Use of the primary literature, and Writing Quality. It includes descriptors for each criteria at four levels of achievement: not addressed, novice, intermediate and proficient (Timmerman et al., 2011). This published rubric has universal applicability in biology subjects because it comprehensively describes the process of science and makes the expectations of critical thinking explicit for students so that they know the goals they are aiming for in their assessment task.

Building capacity in students to self-regulate their learning

The two themes of *guidance and feedback* and *involving students in assessment* are also relevant to solving the problem of building capacity in students to self-regulate their learning. It is critically important for students to be independent, self-regulating learners by the time they graduate because the workplace lacks the scaffolded support for learning that the university environment provides (Boud & Falchikov, 2006). Commencing undergraduate students do not usually have well developed skills in judgment of quality so it is our responsibility as educators to integrate opportunities for development in the curriculum (Boud & Molloy, 2013). In their comparison of models of feedback in biology and engineering, Boud and Molloy (2013) argue that the traditional model, which positions the teacher as the driver and provider of feedback and the student as a dependent receiver, is not sustainable and does not equip students to become independent learners. A sustainable model of feedback places the students at the centre of the learning and develops capacity in students to generate and solicit their own feedback (Boud & Molloy, 2013). This capacity can be developed through assessment practices such as self and peer assessment, which have a central feature of students making judgements about their own and the work of others (Boud & Molloy, 2013).

Hounsell et al.'s (2008) analysis of first and final year bioscience students' perceptions of the guidance and feedback provided on assessments resulted in the development of a guidance-feedback loop framework. The framework identified six interrelated steps and potential trouble-spots where engagement and learning can be compromised and Hounsell et al. (2008) argued that all of the steps outlined in the guidance-feedback loop should be integrated into a single curriculum to maximise the effect on student learning. Orsmond, Maw, Park, Gomez and Crook (2013) proposed the GOALS framework (G = Grasping the objectives or purpose for learning, O = Orientate the student to 'self', A = Actions required to provide dialogue opportunities and enhance self-regulation, L = Learning evaluation opportunities, S = Strategies for moving on) as a tool for implementing the principles of models such as Hounsell et al.'s (2008) guidance-feedback loop, with the aim of facilitating development of students as effective, independent learners. The GOALS framework emphasises peer-discussion, evaluation, self-regulation, the use of exemplars and provides students with opportunities to engage with self- and peer-assessment (Orsmond et al., 2013). This framework builds on a series of papers by Orsmond and colleagues (Orsmond, Merry & Reiling, 2000; Orsmond, Merry & Reiling, 2002; Orsmond, Merry & Callaghan, 2004a; Orsmond, Merry & Reiling, 2004b; Orsmond & Merry, 2013) which discuss how assessment criteria, exemplars, self and peer assessment develop self-regulation skills in students.

One of the key findings from the investigation by Hounsell et al. (2008) was the importance of making assessment expectations clear to students. Bird and Yucel (2013) described an integrated writing programme of tasks for first-year biology students called the Developing Understanding for Assessment (DUAL) programme. The DUAL programme included students marking exemplar reports of varying standards using a rubric which contained statements of assessment criteria and standards, a moderation discussion of the quality of the exemplars with peers and markers, and a peer review of a draft of their first report. A qualitative evaluation of these elements of the programme found strong evidence that the marking and discussion of exemplar reports with peers and demonstrators clarified expectations of scientific report writing for students (Yucel, Bird, Blanksby & Young, 2014). The opportunity for structured discussion about assessment criteria and standards between peers and markers was seen as the most important element of the programme (Yucel et al., 2014). In this discussion, students could clarify explicit statements and develop a tacit knowledge base to enhance their ability to judge the quality of others' work and their own (Yucel et al., 2014). These findings support previous work by Orsmond et al. (2002) which found that the discussion of exemplars and assessment criteria with peers and tutors increased the biology students' understanding of the expectations and resulted in improved student performance.

Peer assessment has been used in the biology discipline in both formative and summative exercises. Liang and Tsai (2010) implemented an online self and peer assessment activity where students submitted a written report on a biology topic of interest three times. On each submission, students received feedback on their written report from other students which they used to make improvements prior to the next submission. The study found that the quality of the scientific writing (with respect to coverage, richness and organisation) improved over time as a result of the peer review (Liang & Tsai, 2010). The students also benefited by participating in multiple rounds of self and peer review where they evaluated each piece of work repeatedly and saw how it changed and improved over time (Liang & Tsai, 2010). It is likely that this experience would have further developed student understanding of assessment criteria and standards and their ability to judge quality. A similar peer review exercise within the DUAL programme was rated by the majority of students as helpful for improving their biology reports (Yucel et al., 2014). The major concern of students was that they received either insufficient or poor quality feedback from their peers (Yucel et al., 2014). This finding reinforced the importance of emphasising that the process of *giving* feedback to peers might benefit a student just as much as *receiving* feedback on their own report (Yucel et al., 2014).

Reliability of peer markers has also been the subject of several studies. Hafner and Hafner (2003) found students marked collaborative oral presentations as consistently as instructors in an evolutionary biology subject, whereas Langan, Wheater, Shaw, Haines, Cullen, Boyle, Penney, Oldekop, Ashcroft, Lockey and Preziosi (2005) found marks awarded by peers were strongly correlated with tutor's marks, but were on average 5% higher. Overall these studies find that learning outcomes are improved when students are involved in the assessment procedure and when understanding of assessment criteria and standards occurs through the development or discussion of criteria prior to peer marking. Orsmond et al. (2004a) found that students were more effective peer reviewers when they discussed the assessment criteria with peers and tutors before participating in the review. Langan et al. (2005) also found that peer markers who had been involved with development of criteria before marking, marked more similarly to tutors compared to peer markers who had not been involved. Despite its obvious benefits, a recent survey of 1535 American academics found that only 41.4% included peer review or assessment exercises in biology subjects (Goubeaud, 2010), indicating that this valuable assessment practice is still under-utilised.

A critical part of effective feedback is clear linking of assessment tasks so that students can apply the feedback, and what they learned, to the next task (Hattie & Timperley, 2007). The DUAL programme described in Bird and Yucel (2013) included a formative task where students completed an action plan which applied the feedback provided on their first report to their second report. A qualitative evaluation of this element of the programme found that the action plan provides an effective tool for students to reflect on the feedback received, decode it and explain how they plan to avoid the same mistakes in their next task (Bird & Yucel, in preparation). The action plan encourages the use of self-assessment skills when students check that errors are not repeated in their subsequent work (Bird & Yucel, in preparation). A study by Orsmond and Merry (2013) investigated student approaches in dealing with tutor-feedback on their work. The study found that high-achieving students actively use feedback by self-assessing their work, seeking clarification for a deeper understanding of the feedback and resolving issues through discussion with peers (Orsmond & Merry, 2013). These attributes align with the goals of sustainable assessment and describe students who will graduate as self-regulated learners. In comparison, non-high achieving students had only a modest capacity for self-assessment and were strongly reliant on the tutor for feedback and strategies for improvement (Orsmond & Merry, 2013). Students with under developed self-assessment skills are the ones with the most to gain from assessment tasks which develop their knowledge and understanding of expectations and standards and their capacity to judge the quality of work (Orsmond & Merry, 2013).

The major problems, related to assessment, and their solutions are summarised in Table 4.

Table 4. A summary of the major problems and solutions identified and discussed in the last decade of assessment in biology literature.

Problem	Solution
Improving the cognitive level of exam and test questions	Tools and strategies for revising exam and quiz questions so that they assess higher order cognitive skills have been trialled and evaluated by biology educators and shown to be effective.
Creating opportunities for formative assessment of content	In-class quizzes, test question design and critique activities and model answer or exemplar exam questions all provide effective feedback on learning of subject material.
Forms of assessment which develop critical thinking skills in students	Critical thinking skills are developed through writing about biology and biological research.
Building capacity in students to self-regulate their learning	Sustainable assessment practices such as applying assessment criteria to exemplars, participating in moderation discussions and peer assessment, self-assessing work and applying feedback effectively to subsequent assessment tasks, develop students' skills in self-regulating learning.

Implications for practice

Our challenge as biology educators is to design assessment practices which support student learning of disciplinary knowledge and skills and also equip students with well-developed critical thinking skills and the capacity to self-regulate their own learning. This review discussed four of the major problems in biology assessment and research conducted over the last decade. There is an on-going need in the discipline of biology for assessment practices which encourage deeper learning of disciplinary knowledge and develop critical thinking skills in our biology students. Happily, many of the innovations discussed and evaluated (e.g., in-class quizzes, design and critique of multiple choice questions, participation in self and peer assessment activities) meet those goals. The majority of research described includes assessment practices *for* rather than *of* learning, where guidance and feedback are integrated into the curriculum (e.g., provision and discussion of assessment criteria and exemplars, self and peer review exercises, sequences of tasks where students have the opportunity to improve over time) and are student rather than teacher focussed (e.g., student-derived assessment criteria, peer assessment). Widespread implementation of these types of assessment activities develops graduates that are knowledgeable, independent learners who are well equipped for the workforce.

Reflecting on our own practice is an excellent first step towards increasing adoption of new assessment practices. We can lead by example within our personal spheres of influence: in the subjects we teach, our disciplinary colleagues who teach in the same or different subjects and our colleagues from other schools, faculties and institutions. Conducting this review inspired me to reflect on the learning and assessment of critical thinking skills in a subject I coordinate. My hope is that this review will provoke others to reflect on their current assessment practices, revise their own curricula, evaluate the impact on student learning and disseminate their findings through their networks of colleagues. A vision, such as the one outlined in the *Vision and Change in undergraduate biology education* statement, where undergraduate biology students will graduate with a '*well-defined level of functional biology literacy and critical thinking skills'* (Brewer & Smith, 2011; p 5), will only be realised if more biology educators rigorously evaluate their curriculum reforms and gather evidence that the new approaches improve learning outcomes. We all need to convince our disciplinary colleagues that change is worthwhile.

Acknowledgments

Many thanks to Ashley Edwards and Pam Hurst for providing feedback on an early draft of this manuscript.

References

Bird, F.L. & Yucel. R. (2013). Improving marking reliability of scientific writing with the developing understanding of assessment for learning programme. *Assessment & Evaluation in Higher Education, 38*(5), 536–553.

Bird, F.L. & Yucel. R. (in preparation). Feedback codes and action plans: building the capacity of first-year students to apply feedback to a scientific report

Bissell, A.N. & Lemons, P.P. (2006). A new method for assessing critical thinking in the classroom. *BioScience*, *56*(1), 66-72.

Bottomley, S. & Denny, P. (2011). A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochemistry and Molecular Biology Education*, *39*(5), 352–361.

Boud, D. & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*, 399–413.

Boud, D. & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, *38*(6), 698-712.

Boud, D. and Associates (2010). *Assessment 2020: Seven propositions for assessment reforming higher education*. Sydney: Australian Learning and Teaching Council. Retrieved September 1, 2013, from: http://www.olt.gov.au/resources/good-practice?text=assessment+2020

Brewer, C. & Smith, D. (2011). *Vision and change in undergraduate biology education. A call to action.* American Academy for the Advancement of Science. Washington, DC. Retrieved September 1, 2013, from: http://visionandchange.org/files/2011/03/Revised-Vision-and-Change-Final-Report.pdf

Clase, K.L., Gundlach, E. & Pelaez, N.J. (2010). Calibrated Peer Review for computer-assisted learning of biological research competencies. *Biochemistry and Molecular Biology Education*, *38*(5), 290–295.

Crowe, A., Dirks, C. & Wenderoth, M.P. (2008). Biology in Bloom: Implementing Bloom's Taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, *7*, 368–381.

Downs, C.T. (2006). What should make up a final mark for a course? An investigation into the academic performance of first year Bioscience students. *Assessment & Evaluation in Higher Education, 31*(3), 345-364.

Dymond, S.J., Scheifele, L.Z., Richardson, S., Lee, P., Chandrasegaran, S., Bader, J.S. & Boeke, J.D. (2009). Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary build-a-genome course. *Genetics, 181*, 13–21.

Ellery, K. (2008). Assessment for learning: a case study using feedback effectively in an essay-style test. *Assessment & Evaluation in Higher Education, 33*(4), 421-429.

Freeman, S., Haak, D. & Wenderoth, M.P. (2011). Increased course structure improves performance in introductory biology. *CBE—Life Sciences Education*, *10*, 175–186.

Gehring, K.M. & Eastman, D.A. (2008). Information fluency for undergraduate biology majors: applications of inquiry-based learning in a developmental biology course. *CBE—Life Sciences Education, 7*, 54–63.

Gibson, L.R. & Liebman, M. (2003). A laboratory exercise for teaching plan interference and relative growth rate concepts. *Weed Technology*, *17*(2), 392-402.

Gliddon, C.M. & Rosengren, R. J. (2012). A laboratory course for teaching laboratory techniques, experimental design, statistical analysis, and peer review process to undergraduate science students. *Biochemistry and Molecular Biology Education, 40*(6), 364–371.

Goubeaud, K. (2010). How is science learning assessed at the postsecondary level? Assessment and grading practices in college biology, chemistry and physics. *Journal of Science Educational Technology, 19*, 237–245.

Gunersel, A.B. & Simpson, N.J. (2010) Instructors' uses, experiences, thoughts and suggestions regarding Calibrated Peer Review. *Assessment & Evaluation in Higher Education, 35*(7), 771-781.

Hafner, J. & Hafner, P. (2003) Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education, 25*(12), 1509-1528.

Haigh, M. (2007). Sustaining learning through assessment: an evaluation of the value of a weekly class quiz. *Assessment & Evaluation in Higher Education, 32*(4), 457-474.

Hattie¸ J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research 77*(1), 81-112.

Hickey, D.T., Taasoobshirazi, G. & Cross, D. (2012). Assessment as learning: enhancing discourse, understanding, and achievement in innovative science curricula. *Journal of Research in Science Teaching, 49*(10), 1240–1270.

Hoskinson, A.M., Caballero, M.D. & Knight, J.K. (2013). How can we improve problem solving in undergraduate biology? Applying lessons from 30 years of physics education research. *CBE—Life Sciences Education, 12*, 153–161.

Hounsell, D., McCune, V., Hounsell, J. & Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research & Development, 27*(1), 55-67.

Huxham, M. (2007). Fast and effective feedback: are model answers the answer? *Assessment & Evaluation in Higher Education, 32*(6), 601-611.

Huxham, M., Campbell, F. & Westwood, J. (2012). Oral versus written assessments: a test of student performance and attitudes. *Assessment & Evaluation in Higher Education, 37*(1), 125-136.

Jones, S., Yates, B. & Kelder J. (2011). Learning and teaching academic standards statement. Australian Learning & Teaching Council. Retrieved September 1, 2013, from: http://www.olt.gov.au/resource-learning-and-teaching-academic-standards-science-2011

Kitchen, E., Bell, J.D., Reeve, S., Sudweeks, R.R. & Bradshaw, W.S. (2003). Teaching cell biology in the large-enrolment classroom: methods to promote analytical thinking and assessment of their effectiveness. *Cell Biology Education, 2*, 180–194.

Langan, A.M., Wheater, C.P., Shaw, E.M., Haines, B.J., Cullen, W.R., Boyle, J.C., Penney, D., Oldekop, J.A., Ashcroft, C., Lockey, L. & Preziosi, R.F. (2005). Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education, 30*(1), 21-34.

Liang, J-C. & Tsai, C.-C. (2010). Learning through science writing via online peer assessment in a college biology course. *Internet and Higher Education 13*, 242–247.

Libarkin, J. & Ording, G. (2012). The utility of writing assignments in undergraduate bioscience. *CBE—Life Sciences Education, 11*, 39–46.

Lingard, J., Minasian-Batmanian, L., Vella, G., Cathers, I. & Gonzalez, C. (2009). Do students with well-aligned perceptions of question difficulty perform better? *Assessment & Evaluation in Higher Education, 34*(6), 603-619.

McCune, V. & Hounsell, D. (2005). The development of students' ways of thinking and practising in three final-year biology courses. *Higher Education, 49*(3), 255-289.

Momsen, J.L. Long, T.M., Wyse, S.A. & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE—Life Sciences Education, 9*, 435–440.

Momsen, J.L., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E. & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE—Life Sciences Education, 12*, 239–249.

Moni, R.W., Moni, K.B., Poronnik, P. & Lluka, L.J. (2007). An e-conference to assess human biology in large, first-year classes. *Biochemistry and Molecular Biology Education, 35*(4), 255–262.

Mulnix, A.B. (2003). Investigations of protein structure and function using the scientific literature: an assignment for an undergraduate cell physiology course. *Cell Biology Education, 2*, 248-255.

Offerdahl, E.G. & Tomanek, D. (2011). Changes in instructors' assessment thinking related to experimentation with new strategies, *Assessment & Evaluation in Higher Education, 36*(7), 781-795.

Orsmond, P., Merry, S., Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 25*(1), 23-38.

Orsmond, P., Merry, S. & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 27*(4), 309-323.

Orsmond, P., Merry, S. & Callaghan, A. (2004a). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International, 41*(3), 273-290.

Orsmond, P., Merry, S. & Reiling, K. (2004b). Undergraduate project work: can directed tutor support enhance skills development? *Assessment & Evaluation in Higher Education, 29*(5), 625-642.

Orsmond, P. & Merry, S. (2013). The importance of self-assessment in students' use of tutors' feedback: a qualitative study of high and non-high achieving biology undergraduates. *Assessment & Evaluation in Higher Education, 38*(6), 737-753.

Orsmond, P., Maw, S.J., Park, J.R., Gomez, S. & Crook, A.C. (2013). Moving feedback forward: theory to practice. *Assessment & Evaluation in Higher Education, 38*(2), 240-252.

Peat, M. & Franklin, S. (2002). Supporting student learning: the use of computer–based formative assessment modules. *British Journal of Educational Technology, 33*(5), 515–523.

Preszler, R.W., Dawe, A., Shuster, C. & Shuster, M. (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE—Life Sciences Education, 6*, 29–41.

Quitadamo, I.J. & Kurtz, M.J. (2007). Learning to improve: using writing to increase critical thinking performance in general education biology. *CBE—Life Sciences Education, 6*, 140–154.

Quitadamo, I.J., Faiola, C.L., Johnson, J.E. & Kurtz, M.J. (2008). Community-based inquiry improves critical thinking in general education biology. *CBE—Life Sciences Education, 7*, 327–337.

Ramsden, P. (2003). *Learning to teach in higher education*. 2nd edition. London, New York: Routledge Farmer.

Rebello, C.M., Siegel, M.A., Freyermuth, S.K., Witzig, S.B. & Izci, K. (2012). Development of embedded assessments for learning in biotechnology. *Biochemistry and Molecular Biology Education, 40*(2), 82–88.

Reynolds, J.A. & Thompson, R.J. Jr. (2011). Want to improve undergraduate thesis writing? Engage students and their faculty readers in scientific peer review. *CBE—Life Sciences Education, 10*, 209–215.

Richmond, G., Merritt, B., Urban-Lurain, M. & Parke, J. (2010). The development of a conceptual framework and tools to assess undergraduates' principled use of models in cellular biology. *CBE—Life Sciences Education, 9*, 441–452.

Scoles, J., Huxham, M. & McArthur, J. (2013). No longer exempt from good practice: using exemplars to close the feedback gap for exams, *Assessment & Evaluation in Higher Education, 38*(6), 631-645,

Stanger-Hall, K.F. (2012). Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE—Life Sciences Education, 11*, 294–306.

Sundberg, M.D. (2002). Assessing student learning. *Cell Biology Education, 1*, 11–15.

Timmerman, B.E.C., Strickland, D.E., Johnson, R.L. & Payne, J.R. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education, 36*(5), 509-547.

Ueckert, C., Adams, A. & Lock, J. (2011). Redesigning a large-enrolment introductory biology course. *CBE—Life Sciences Education, 10*, 164–174.

Yucel, R., Bird, F.L., Blanksby, T. & Young, J. (2014). The road to self-assessment: exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2014.880400

Wood, W.B. (2009). Innovations in teaching undergraduate biology and why we need them. *Annual Review of Cell and Developmental Biology, 25*, 93-112.

Appendix I. List of topics and references grouped under each assessment theme

Theme
1. What guidance and feedback is given, by what means and when • Models or frameworks for improving learning from feedback (Hounsell et al., 2008; Orsmond et al., 2013) • Student use feedback in different ways (Orsmond & Merry, 2013) • Quality of feedback is perceived as important by students (McCune & Hounsell, 2005) • Guidance on posters (Orsmond et al., 2002), exam questions (Huxham, 2007; Scoles et al., 2013), laboratory reports (Bird & Yucel, 2013) provided in form of exemplars and model answers • Peer review of laboratory report (Gibson & Liebman, 2003; Liang & Tsai, 2010; Bird & Yucel 2013), theses (Reynolds & Thompson, 2011) and poster drafts (Orsmond et al., 2004a) before summative assessment • Peer review of multiple choice questions (Bottomley & Denny, 2011) • Tutor feedback on draft of poster, essay (Quitadamo & Kurtz, 2007), research proposals (Quitadamo et al., 2008), scientific manuscript (Gliddon & Rosengren, 2012), quality of a scientific argument (Hickey et al., 2012) before summative assessment • A series of assessments allowing feedback to be applied to subsequent tasks (Ellery, 2008; Gehring & Eastman, 2008; Libarkin & Ording, 2012) • Mastery of laboratory skills by practising and adjusting technique in response to feedback from tutors • Feedback on knowledge and understanding of content via in-class (Kitchen et al., 2003; Preszler et al., 2007; Wood, 2009; Freeman et al., 2011; Ueckert et al., 2011) and online quizzes (Peat & Franklin, 2002)
2. Student involvement in assessment • Design multiple choice questions (Bottomley & Denny, 2011) • Student-derived assessment criteria (Orsmond, 2000; Langan et al., 2005) • Self-assessment of reports (Liang & Tsai 2010), posters (Orsmond et al., 2002), theses (Reynolds & Thompson, 2011) • Peer assessment of poster and oral presentations (Orsmond et al., 2002; Hafner & Hafner, 2003; Langan et al., 2005), group work (Mulnix, 2003; Quitadamo et al., 2008), reports (Liang & Tsai, 2010) • Marking exemplars (Bird & Yucel, 2013) • Discussion of criteria and exemplars with peers and tutors (Orsmond et al., 2004a, Reynolds & Thompson, 2011; Bird & Yucel, 2013) • Sustainable assessment practices develop students as assessors (Boud & Molloy, 2013)
3. Modes of assessment and balance between them • Modes of assessment used in biology compared to other disciplines (Goubeaud, 2010) • Adopting problem-solving approaches, commonly used in physics, for biology (Hoskinson et al., 2013) • Oral versus written examinations (Huxham et al., 2012) • Analysis of student performance on different assessment tasks (Downs, 2006) • Written versus quiz assessment of learning from laboratory exercises (Quitadamo & Kurtz, 2007)

	• Balancing constructed response and multiple-choice questions in exams (Stanger-Hall, 2012) • Correlating in-class quiz performance with performance on different assessment tasks (Haigh, 2007)
4.	**Criteria used and their relative weighting** • Use of rubrics to assess content and critical thinking in constructed response test questions (Bissel & Lemons, 2006) • Evaluation of a rubric for peer assessment of oral presentations (Hafner & Hafner, 2003), scientific discourse (Hickey et al., 2012), critical thinking (Timmerman et al., 2011) • Student-derived assessment criteria (Orsmond, 2000; Langan et al., 2005)
5.	**Perceptions and experiences of staff** • Staff perceptions of Calibrated Peer Review (Gunersel & Simpson, 2010) • Student perceptions of their teaching & learning environments (McCune & Hounsell, 2005) • Student perceptions of feedback (Hounsell et al., 2008) • Staff and student perceptions of meaning of assessment criteria (Orsmond et al., 2004) and the difficulty of assessment items (Lingard et al., 2009) • Effect of trialling new formative tasks on the way staff think about assessment practices (Offerdahl & Tomanek, 2011)
6.	**Enhancing assessment practices** • Analysis of the cognitive level of assessments such as multiple choice questions (Momsen et al., 2010) • Analysis of assessment practices being utilised and their impact on learning (Momsen et al., 2013) • Designing multiple choice (Crowe et al., 2008) and constructed response questions (Bissell & Lemons, 2006) to test higher order thinking skills • Designing questions about biological models (Richmond et al., 2010) • Design principles for embedded assessments in inquiry based learning (Rebello et al., 2012)
7.	**Use of new technology in assessment** • Design and critique multiple choice questions using PeerWise (Bottomley & Denny, 2011) • Conduct peer reviews using Calibrated Peer Review (Clase et al., 2010; Gunersel & Simpson, 2010) or other online tools (Liang & Tsai, 2010) • Student poster presentations at an e-conference (Moni et al., 2007) • Knowledge and understanding of content tested using classroom response 'clicker' systems (Kitchen et al., 2003; Preszler et al., 2007; Wood, 2009; Freeman et al., 2011; Ueckert et al., 2011) and online quizzes (Peat & Franklin, 2002)