J Sci Educ Technol (2016) 25:203–221
DOI 10.1007/s10956-015-9586-0

Technology-Enhanced Formative Assessment of Plant Identification

Ricardo Conejo[1] · Juan Ignacio Garcia-Viñas[2] · Aitor Gastón[2] · Beatriz Barros[1]

Published online: 28 October 2015

Abstract Developing plant identification skills is an important part of the curriculum of any botany course in higher education. Frequent practice with dried and fresh plants is necessary to recognize the diversity of forms, states, and details that a species can present. We have developed a web-based assessment system for mobile devices that is able to pose appropriate questions according to the location of the student. A student's location can be obtained using the device position or by scanning a QR code attached to a dried plant sheet in a herbarium or to a fresh plant in an arboretum. The assessment questions are complemented with elaborated feedback that, according to the students' responses, provides indications of possible mistakes and correct answers. Three experiments were designed to measure the effectiveness of the formative assessment using dried and fresh plants. Three questionnaires were used to evaluate the system performance from the students' perspective. The results clearly indicate that formative assessment is objectively effective compared to traditional methods and that the students' attitudes towards the system were very positive.

Keywords Field-based learning · Web-based assessment · Formative assessment · Educational technology · Plant identification

✉ Ricardo Conejo
conejo@lcc.uma.es

1 E.T.S.Ing. Informática, Universidad de Malaga, 29071 Málaga, Spain

2 E.T. Ing. Forestal, Universidad Politecnica de Madrid, 28071 Madrid, Spain

Introduction

An important objective of the botany course at the *Escuela de Ingeniería de Montes, Forestal y del Medio Natural* (Forestry and Natural Environment Engineering School) of the Technical University of Madrid is to instruct students in the identification of approximately 150 species from their geographic area. This objective may have a forestry application. Although the identification of a species is commonly based on morphological characteristics, which are taught using dichotomous keys, a species is also associated with habitat, soil conditions, climatic type, accompanying species, and other factors. The recognition process is complex due to the seasonal variability of species. To develop identification skills, students need frequent practice to recognize the diversity of forms, states, and details that species can present.

Identification skills are taught and learned in a laboratory that uses dried plants from a herbarium complemented by photographs and field visits to different locations for fresh plant identification. The classes with dried plants in the laboratory have the advantage of previous preparation and that the study is not restricted to any geographical area. Plants from different locations can be presented together according to morphological or taxonomical similarities. However, dried plants have the disadvantage of changes in their colour and general appearance and often in their phenological state (flowers and fruits). Thus, students are limited by interacting with just a fragment of the plant. Photographs complement the comprehension of the dried plants, but they are also restricted to showing just a few details. Students find it difficult to synthesize the knowledge from these elements that allows recognition. Field visits are very important for learning in life and environmental sciences (Scott et al. 2012), and allow students to have an experience that is closer to professional practice. However, a limiting factor is the number of practical classes and field visits that can be offered for a given taxonomical group. Thus, students are encouraged to engage in self-learning using dichotomous keys and other material.

Plant identification skills require different types of knowledge. It is important to note that merely recalling theoretical knowledge is insufficient to recognize a plant. Students should know where to look and how to determine the characteristic elements that differentiate a species. Moreover, recognizing pre-processed dried plants does not automatically imply recognition of a fresh specimen in the field. In many cases, some important classification elements are missing (for example, a plant may have no flowers, or the specimen may be young and have a different appearance). However, other elements can help to distinguish and classify a specimen, such as location, soil, climatic type, and accompanying species. According to Bloom's taxonomy (Bloom et al. 1956; Anderson et al. 2001), the identification of plants in a real-case environment requires at least three initial levels of knowledge, comprehension, and application, and in some cases the ability to analyse all of the data that are collected. In summary, the identification of live plant species in the field requires high-order cognitive skills (HOCS) as described by Zoller and Tsaparlis (1997) and Crowe et al. (2008). A recent survey on the application of technology to biological education (Lee and Tsai 2013) suggested that more studies should explore the role of technologies and associated pedagogies in fostering high-order skills.

Limited research exists on the most effective methods for teaching species identification. Ohkawa (2000) proposed the use of synoptical keys for junior and senior high-school students. Randler (2008) focused on primary school and the identification of animals. He concludes that black and white dichotomous keys are a good alternative and mentioned that outdoor education should be enhanced by prior learning within the classroom. Uno (2009) described the current situation of botanical education in the US and claimed that students should be provided with ample opportunities to engage in the processes of discovery in class and to apply their conceptual understanding and the processes of science to their lives outside class. Silva et al. (2011) proposed a computer-based interactive dichotomous key to support the identification of fresh plants for higher education students. Stagg and Donkin (2013) described a pilot study that used dichotomous keys, a word-association exercise based on a mnemonic approach, and a pictorial card game. The main problem of plant identification practice, either using a notebook or using a computer-supported dichotomous key, is that an expert teacher has to be available to confirm that the identification was correct. This fact implies practicing in groups with the teacher or practicing alone and receives the results and explanation later.

It is well known that knowledge is constructed primarily by means of student activities (Biggs 2011). According to constructivism theories, interactive activities in which learners play active roles are more effective than activities in which learners remain passive (Huang 2002). Gardner and Belland (2012) drew attention to the importance of active learning in biology instruction and proposed technology-enhanced activities, among others. On the other hand, Klionsky (2008) summarized some findings in cognitive psychology that indicate that repeated studying turns out to be relatively ineffective in enhancing learning, whereas testing after studying has a beneficial effect (McDaniel et al. 2007; Karpicke and Roediger 2008).

Formative assessment is described by Black and Wiliam (2009) as the use of assessment by teachers for learning practices to improve student achievement. This concept is also known as *assessment for learning* (Cooper and Cowie 2010). There are many examples of formative assessment in psychology (Buchanan 2000), computer science (Guzman et al. 2007), physics (Dufresne and Gerase 2004), chemistry (Momsen et al. 2013), and biology (Preszler et al. 2007; Crowe et al. 2008). Ongoing authentic assessment activities and interactive formative feedback have been identified as being important characteristics that can address threats to validity and reliability within the context of online formative assessment (Gikandi et al. 2011).

Feedback has been viewed as a primary strategy in *formative assessment*. Feedback is defined as any piece of information that is given to a student after a student's action. An important determinant of the effectiveness of formative assessment is the quality of the feedback received by learners. In a testing environment, feedback can be classified according to time and content (Shute 2008). Timing feedback may take the form of: (1) *immediate feedback*: questions are posed one by one, and feedback is provided after each answer is given; (2) *delayed feedback*: there are many variants of delayed feedback, but it mainly refers to feedback given at the end of a test or afterwards. Delayed feedback can take several forms: (1) *correct response* (usually known as KR): feedback only consists in stating that the given response was right or wrong; (2) *given answer* (usually known as KCR): feedback confirms a correct response, but also provides the right answer in case of a wrong response; and (3) *elaborate feedback* (known as KCR + EL): feedback also includes further detailed explanations. A recent study on feedback in questionnaires (Fabienne et al. 2012) suggested that students pay more attention to immediate feedback than to delayed feedback. Furthermore, although KCR and

KCR + EL are clearly superior to KR, it is not possible to differentiate between KCR and KCR + EL. Jordan (2012) also suggested that feedback is most effective when it can be understood by the student, when it is tailored to their mistakes, and when the students are prompted rather than being given the answer.

Computer technology represents a powerful tool for assessment design and delivery, thus giving rise to the new fields of computer-based assessment (CBA), computer-based testing (CBT), and, more recently, web-based assessment (WBA). Test delivery, automatic grading, and immediate feedback are clear advantages. Technology facilitates the creation and maintenance of large item banks that can be reused and can help to promote active learning and continuous *formative assessment*. The actual challenge is to construct item banks that automatically assess HOCS (Zoller 2001; Crowe et al. 2008) while avoiding the problems of multiple-choice questions (Palmer and Devitt 2007; Stanger-Hall 2012; Tractenberg et al. 2013).

However, CBA is still an open and relatively unexplored field. Many teachers who use CBA have only explored the features that can also be performed by paper-and-pencil assessment. The field is rapidly evolving with the arrival of each technological innovation.

Mobile devices, such as smartphones and tablets, are changing the way in which we access Internet applications and opening new possibilities in the field of CBA. Using the communication and computing capabilities of these devices, tests can be delivered not only in the classroom but also in an open environment in which students walk around. The selection of items to be delivered can be adapted to a student's location and circumstances. This technology allows flexible indoor testing and even outdoor testing.

We propose a system for the automatic formative assessment of plant recognition in two different scenarios: (1) dried plants in a laboratory; and (2) fresh plants. In both cases, students can receive elaborated feedback (KCR + EL) on their responses.

Assessment is implemented using a WBA tool called Siette (Conejo et al. 2004). The system supports multiple-choice questions (MCQ) and constructed responses (CR) that are corrected according to a regular expression (see "Geolocalized Questions and QR Codes" section). Tests are delivered through a web interface that can be used in desktop computers, smartphones, and tablets. Using these devices, the system is able to adapt questions to the location of the user by using GPS, Wi-Fi, or 3G network signals. This information allows questions to be posed that are related to a given location of a fresh plant. It is also possible to directly select the question to be posed by scanning a bidimensional QR code.[1] This feature allows attaching QR codes to dried plant sheets in a herbarium or to fresh plants in an arboretum. Cameras that are embedded in the mobile devices and equipped with suitable software that can scan QR codes identify them as an Internet address (URL) and automatically display the corresponding question in a web browser (see "Test Delivery", "Identification of Dried Plants: Measuring Learning Gains", and "Identification of Fresh Plants: Measuring the Effect of Technology-Enhanced Assessment" sections).

Bidimensional codes have been used in educational applications in the context of pervasive or mobile learning. Summaries of such applications can be found in Law and So (2010), Ozcelik and Acarturk (2011), Uluyol and Agca (2012), and Lucke and Rensing (2014). Some of these applications are direct extensions of well-known ideas: for example, interactively displaying the results of catalogue searches at a library or labelling student assignment submission sheets. A very interesting development is the integration of printed and on-line material that can be accessed through QR codes and smartphones (Uluyol and Agca 2012; Huang et al. 2012). Some applications have been documented that use bidimensional codes for outdoor learning. One such is the Myst game platform (Laine et al. 2010) in which players have to solve a set of enigmas. The game area is divided into several subareas, and each subarea has a unique set of enigmas to which it is related. 2D barcodes are used to provide context-awareness to the game so that enigmas and hints are presented in just the right subarea. Mobile and web-based testing has been described as a highly motivating activity for students (Romero et al. 2009), even though their system does not include location-awareness. Hwang and Chang (2011) described an on-site formative assessment experience based on mobile devices and obtained positive results. Santos et al. (2014) described an interesting system called QuestInSitu for geolocalized testing with MCQ that only uses GPS location. Our proposal differs from the previous studies in several ways: (1) it combines geolocation based on 3G and Wi-Fi signals with QR codes, thereby increasing location accuracy and allowing indoor and outdoor assessments; (2) it uses the rich question format, including constructed response with short answers; and (3) it includes adaptive elaborated feedback.

The final aim of these experiments was to evaluate whether the application of this technology helps students to learn and motivates them to practice more. To evaluate the performance of the system and the effects of formative assessment, three different experiments were performed:

[1] There are several matrix codes (Kato and Tan 2007). Among them, the QR code is perhaps the most common in mobile applications, because it has an official standard (ISO 2006) and can be used freely. QR code (quick response code) is a trademark of the Japanese company Denso Wave (a Toyota subsidiary). It can encode all types of data. In particular, it can encode a string and hence a URL.

(1) the evaluation of system performance and the usability of QR codes attached to fresh plants in the arboretum. We were also interested in the comparative assessment of fresh plants and photographic identification skills; (2) the evaluation of the effectiveness of formative assessment as a learning tool for the identification of dried plants labelled with QR codes; and (3) the comparison of technology-enhanced formative assessment with traditional plant identification training sessions based on field notebooks.

Uno (2009) proposed certain principles that should be taken into account for the effective learning of botany (in fact they are applicable to all scientific areas). The final aim of our assessment system was to fulfil at least three of these principles: (a) "*Learners have different strategies, approaches, abilities, and learning styles that are a function of the interaction between their heredity and their prior experiences*". Using a system of self-regulated formative assessment, students would be allowed to design and conduct their own learning experiences. They would be able to select the way in which they want to practice identification skills; (b) "*Learners' motivation to learn and their sense of self affect what is learned, how much is learned, and how much effort will be put into the learning process*". Continuous grading is a direct indicator of how much has been learned. Moreover, students have clearly expressed that the system increased their motivation; and (c) "*The practices and activities in which people engage while learning shapes what is learned*". Formative assessment in a real-world situation with fresh and dried plants engages our students and helps learning. The opinions that were expressed in study questionnaires clearly support this assertion. The analysis of the qualitative results further confirms this assertion.

This paper is structured as follows: the next section describes the web-based assessment system and its extension to support context-aware questions. A description of the botany course item bank is also included. "The Experiments" section describes the three experiments. The methodology, results, and specific conclusions of each experiment are analysed in different subsections. Finally, "Conclusions" section summarizes the main conclusions, discusses some limitations, and proposes future directions for this research.

The Assessment System

Siette (Conejo et al. 2004; Guzman et al. 2007) is a domain-independent web-based platform that supports the whole assessment life cycle. It is based on the development of a reusable item bank. Siette allows for the use of different types of questions and supports different question selection criteria, different scoring procedures, etc. Siette also includes an authoring tool and a set of options to analyse test results and generate statistical information. Siette is available at http://www.siette.org.

There are three basic types of questions: multiple choice, single answer (MC-SA); multiple choice, multiple answer (MC-MA); and the constructed response (CR) of short answers. The latter are based on recognizing the student's response by means of a regular expression, that is, a pattern that describes all possible answers. For example, the pattern: "(*P.*|*Pinus*) *nigra* {{subsp|ssp}. *salzmannii*|. *salzmannii*}" accepts "*Pinus nigra*", "*P. nigra salzmanii*", "*Pinus nigra* subsp. *salzmannii*" (and others) as correct answers. Additional patterns can be provided to recognize common mistakes by students. Thus, Siette recognizes a wrong answer and is able to provide appropriate feedback. Feedback is associated with each possible correct or incorrect answer, and includes the correct answer and some explanation or complementary material (KCR + EL according to Shute's (2008) classification). Figure 1 shows examples of some questions and their feedback.

Geolocalized Questions and QR Codes

Any type of question that is supported by Siette and the device's web browser can be attached with a geolocalization attribute or a QR code. Figure 2 shows a partial view of the Siette authoring tool. The stem, responses, feedback, and attributes of the question are edited in different tabs (Fig. 2-1). New fields have been added to the location information in the "Selection" tab. These fields require entering the exact latitude and longitude of the point where the question can be posed. Of course, these numerical data are not easy to determine, and it would be a burden for teachers to enter them directly. To facilitate the work, a pop-up framework has been defined (Fig. 2-2). This connects to the Google Maps API and allows visual identification of the location. Geographical coordinates can be obtained by clicking on a point or entering an address and searching for it. Authors also select the mode in which the question can be accessed: by GPS/Wifi/3G location, by QR code, or both. The GPS mode means that the question will be triggered when the student reaches the location (within a defined circle), while the QR mode means that the question can only be triggered by capturing the corresponding code. However, note that questions with QR codes also have a location attribute, whose use is explained in the next section. If the QR mode is selected, clicking on the QR button will open a pop-up window with the QR code assigned to the question (Fig. 2-3).

The Siette editor includes a tool for searching all of the questions in a certain area and generating a sheet that contains all of the QR codes to make things easier for the

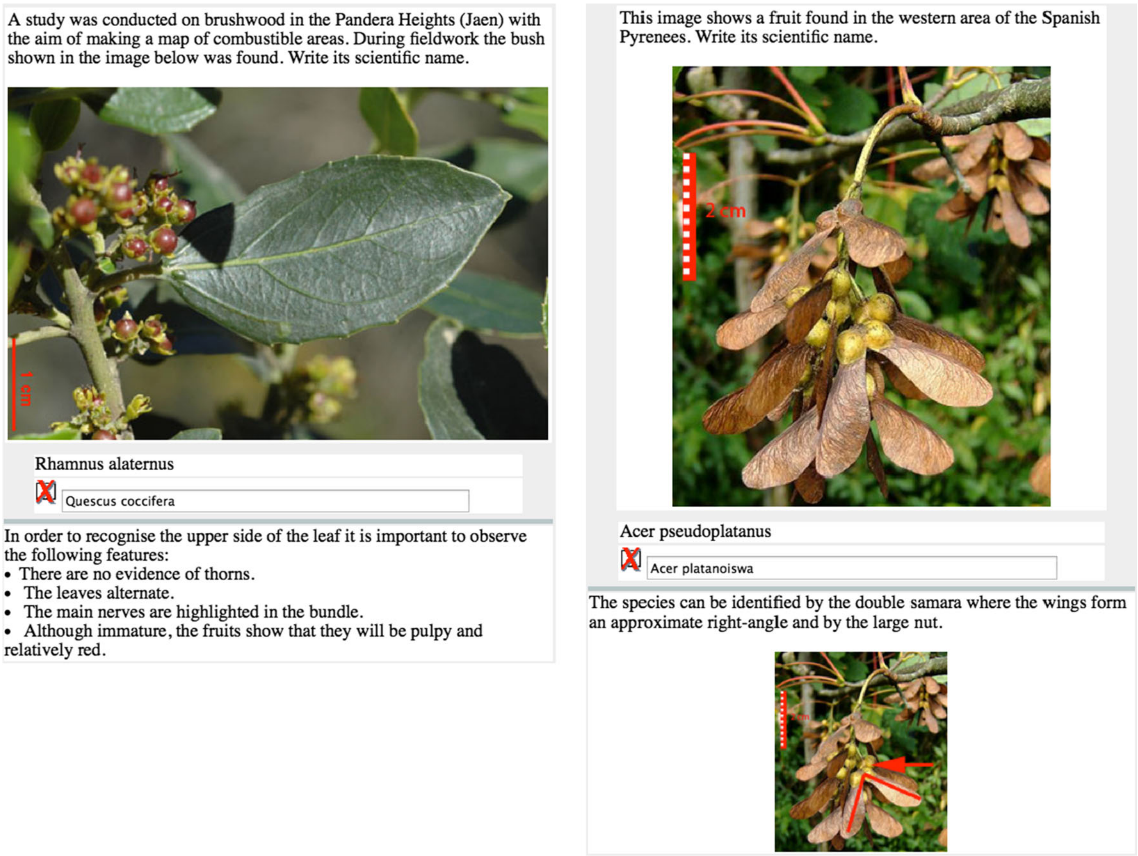


Fig. 1 Two examples of feedback given in response to students' wrong answers

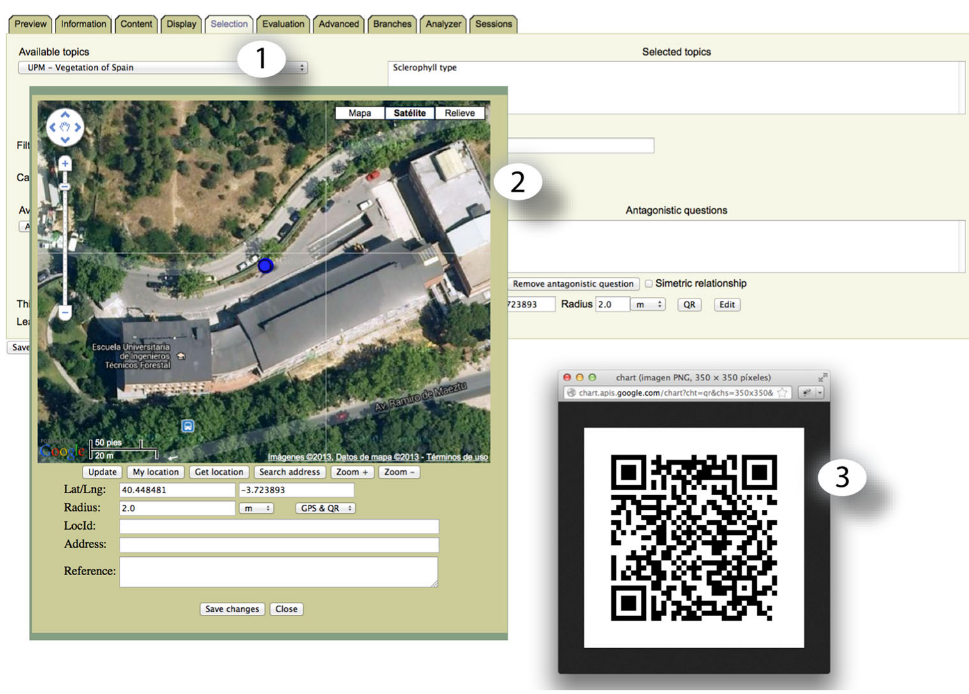


Fig. 2 SIETTE authoring tool. Pop-up frames for location and QR code

teacher. Additionally, the teacher can create a QR code for test initialization (this feature can be useful if the same questions are used in different tests) and generate QR codes for questions without assigning them to any particular test.

Test Delivery

Tests that contain geolocalized or QR-code questions are accessed in the same way as any other tests in the Siette system. First, the user has to login, select a subject, and choose which test to take. A QR code can be read using any scanner application in the mobile device; then the Siette session begins. Direct access to a given test can be granted using a QR code for that test, or it can be included in the QR code of the questions. In this case, the first question that is scanned will request a previous login.

When a test starts, Siette selects the question to pose according to the selection criteria that have been defined. Questions that have a localization attribute are excluded from the item pool. If no questions are available (except those with a localization attribute or a QR code), a *waiting page* is generated and presented to the user. The *waiting page* shows a map that indicates the location of the questions that have not yet been posed. Each location may have a brief description attached that is shown when the cursor is moved across it. These tips assist the user in finding the correct location (Fig. 3). The student should move to any of these locations to trigger a question. Question order is not relevant. When the student reaches the location or scans the QR code of a question (depending on the predefined mode that has been previously declared), the system selects that question and poses it to the student. Obviously, if QR codes are used, the labels that contain the QR codes should have been set previously by the teacher in the appropriate location. Purely geolocalized questions do not require this step, but a tolerance radius has to be defined to address the errors in the estimation of the student's current position.

When a question is posed, the behaviour is the same as a normal test. The question is displayed in the student's mobile device. The student answers the question, and his or her response is returned to the system. At that point, there might be other questions that are attached to the same location or to the same QR code. If this is the case, the system selects one of the questions using the selection criteria. However, if no question is available, a new waiting page is generated that contains the location of the remaining questions. The process continues until the finalization criteria are met.

The Experiments

Location-aware testing was implemented in 2012. At the end of the course, a pilot study experiment was designed to evaluate system performance and compare the results of the assessment using photographs and fresh plants. This experiment is described in "Identification of Fresh Plants: System Usability" section.

A second experiment was designed at the end of the 2013 semester to measure the effectiveness of formative assessment using dried plants. The experiment had a similar design to that of McDaniel et al. (2007). A set of dried plants was studied, some of which had a QR code attached

Fig. 3 Waiting page showing the location of questions

that triggered some reflective questions and provided adaptive feedback in case of error. The hypothesis was that these plants are better recalled than the others. This experiment is described in "Identification of Dried Plants: Measuring Learning Gains" section.

A third experiment was designed to explore the new educational opportunities that this technology provides: to compare it with the existing low-tech alternative of putting numbers on live plant specimens and having students write down the scientific names with pencil and paper; and to measure any educational benefit above what is provided by existing teaching tools. This experiment is described in "Identification of Fresh Plants: Measuring the Effect of Technology-Enhanced Assessment" section.

All of the experiments involved one or more test. To increase the students' motivation, they were told that the score they obtained would be used as part of the continuous evaluation of the course. However, there were many other evaluations, and so the score that was obtained did not represent a high percentage of the final score.

Identification of Fresh Plants: System Usability

The objectives of this experiment were twofold: (1) to evaluate the system's performance in a real-case application and to collect student perceptions and opinions; and (2) to compare plant identification based on photographs with plant identification based on fresh plants. The main research questions were: *Does the system work in a real-case situation? Do students find it to be useful? Is it more difficult for students to identify fresh plants than photographs? Does it take longer to take a field test than a classic test?*

The experiment included 28 volunteer students of the *Escuela de Ingeniería de Montes, Forestal y del Medio Natural* (Forestry and Natural Environment Engineering School) of the Technical University of Madrid. The experiment took place at the end of the semester, when it was assumed that all of the students had mastered the material. The experiment consisted of two tests:

- T_1 A classic test that was delivered with Siette, containing questions about plant recognition based on high-definition photographs
- T_2 A location-aware test, using QR codes that were attached to living specimens of plants at the arboretum of the Forestry School

Each test contained ten questions that were randomly selected out of a pool of 20. The final score is the percentage of correct answers. This randomized selection was designed to avoid cheating and intra-subject communication, especially in the location-aware test that was taken in an open environment. All participants in this experiment took both tests. The classic test (T_1) was taken in a computer room. The location-aware test (T_2) was taken independently by four students at a time, using the four available iPads. The classic test was taken first, but the correct answers and the feedback were not shown until the end of the second test to avoid interference between the tests. In all cases, the expected answer was the scientific name of a plant. In the classic test, the questions included a brief description of a plant's appearance and habitat (see Fig. 4-left). In the location-aware test, a QR code for each question was located beside a fresh plant, and the stem only had the habitat description attached, but not the written description of the plant or its photograph (see Fig. 4-right).

When the students began to take T_2, the system displayed a map with question locations (Fig. 3). They initially walked in the direction that was indicated by the teacher, who made sure they were spread out to prevent them from communicating with each other. Whenever a student found a QR code attached to a plant, he or she scanned it using an iPad application, and the question appeared on the iPad screen. Then the student answered the question, and a new map was displayed containing the location of the remaining questions. The student continued until reaching the maximum number of questions. Figure 5 shows some students taking the location-aware test.

Table 1 shows the test results. A paired *t* test showed that the T_2 score was significant higher than the T_1 score ($p < 0.0001$). Almost all the students increased their scores. The duration of T_2 was slightly longer than T_1 although without reaching statistical significance; this is unsurprising given the time spent on scanning the codes and moving from one site to another. Pearson correlation between tests was low, 0.28, which is mainly explained by the differences in the scores of the low-level students. On the other hand, the correlation between time spent on T_1 and T_2 was high (0.76).

One of the aims of this experiment was to evaluate the system's performance in a real-case application and to determine the users' opinion. First, the system worked correctly because all of the students were able to complete the test in a reasonable time. A survey was conducted to obtain the users' opinion. Table 2 shows the results of the survey, which was anonymous and non-compulsory. In total, 22 of the students answered the survey. Some important questions were asked twice, changing the form of expression to check for consistency. All of the questions were answered using a 1–5 Likert scale where, according to the question, 1 means "*none*", "*very low*", or "*completely disagree*", and 5 means "*all*", "*very much*", or "*completely agree*". The results are given with the mean and standard deviations.

The overall evaluation of the useability of the system and the usefulness of the activity was very positive. All of

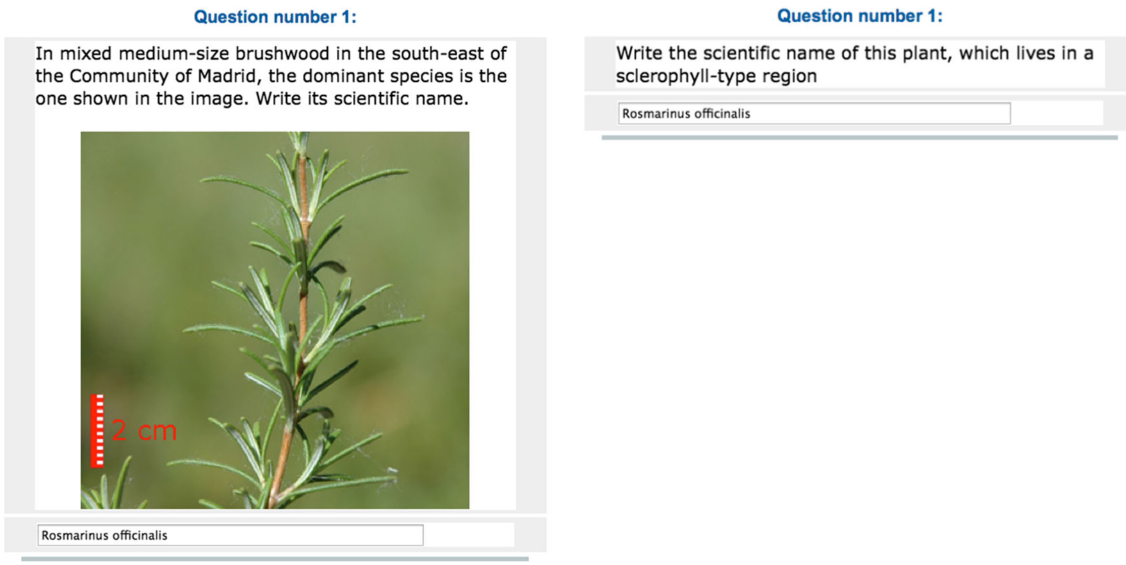


Fig. 4 Questions for T_1 (*left*) and T_2 (*right*). In both cases, the correct answer is Rosmarinus officinalis

Fig. 5 Students taking a field test using QR codes

Table 1 Percentage of correct identifications in T_1 (photographs) and T_2 (fresh plants)

	#Species	#Students	#Cases	#Ident	Score	Time
T_1 (photographs)	10	28	280	123	0.475 (0.499)	10.29 (4.10)
T_2 (fresh plants)	10	28	280	174	0.621 (0.485)	11.28 (4.92)

the participants responded to Question-18 with a value above 4 on the 1–5 Likert scale, with an average of 4.36. This result is especially significant if we take into account that some of the students claimed no previous use of iPads or smartphones. Another interesting result is that almost all the students completely agreed that doing the test with living plants was better than with the photographs (Question-8, 4.86 on the 1–5 Likert scale) and that this type of assessment was appropriate both for formative assessment (Question-16, 4.73) and for summative assessment (Question-17, 4.50).

The end of the questionnaire contained a section for open comments from the students. Most of the comments were positive and suggested different improvements, such as providing hints and feedback to the questions (these features were disabled for the experiment because we wanted to study a pure assessment of these conditions compared to the photographs assessment), the possibility of going back and forth when answering questions, including an optimal path in the map, and even distributing the QR codes in all of the city's gardens and parks. Allowing going back and forth might be a new feature, but it is doubtful that it would be practical in the case of geographically distributed questions.

On the other hand, some students noted that the Internet connection was slow and that the maps took a while to be downloaded (someone also proposed that printed maps be distributed as an alternative). Some complained about the

Table 2 Results of the anonymous questionnaire after the field test

Question	1	2	3	4	5	Mean (SD)
About the student						
1. My experience of using an iPad is…	5	3	7	3	4	2.91 (1.41)
2. My experience with mobile devices is…	0	2	8	5	7	3.77 (1.02)
3. My experience with computer-based evaluation systems is …	1	6	6	9	0	3.05 (0.95)
4. I think that the use of information technologies for learning is helpful	0	2	4	8	8	4.00 (0.38)
About the activity						
5. I enjoyed taking the location-aware test	0	0	1	11	10	4.41 (0.39)
6. It required an extra effort to do it	3	6	8	5	0	2.68 (0.99)
7. I enjoyed this activity	0	0	4	10	8	4.18 (0.73)
8. Doing the test with living plants is better than with photographs	0	0	1	1	20	4.86 (0.47)
About the system						
9. The system worked fine. I was able to complete the test and the system always behaved as expected	0	1	6	8	7	3.95 (0.90)
10. The system interface is intuitive. I always found the buttons and options whenever I needed them	0	2	4	8	8	4.00 (0.98)
11. It was easy for me to locate the questions in the Arboretum	0	1	1	9	11	4.36 (0.79)
12. The annotations attached to each question in the map were useful	3	6	7	3	3	2.86 (1.25)
13. At any time I knew where I was and what was left to finish the test	0	2	2	5	13	4.32 (0.99)
14. The system is easy to use	0	0	1	9	12	4.50 (0.60)
15. I prefer to take the test with the iPad than with the computer	0	0	1	3	18	4.77 (0.53)
16. This type of test is appropriate for practicing	0	0	2	2	18	4.73 (0.63)
17. This type of test is appropriate to assess student knowledge	0	0	3	5	14	4.50 (0.74)
About the system						
18. Overall evaluation of the use of the system and the activity	0	0	1	12	9	4.36 (0.58)

scarcity of mobile devices and the extra time that was spent on taking the test.

Identification of Dried Plants: Measuring Learning Gains

The aim of this experiment was to validate the hypothesis that formative assessment on studied material promotes subsequent learning and retention of that material. The specific research questions were: *Do students learn better to identify species from dried plants on herbarium sheets if they take a formative assessment at the end of the class? Is this effect related to the students' knowledge levels?* A parallel objective was to determine the students' subjective perceptions of the effect of the formative assessment and to collect their opinions about the usability of the system.

The experiment included 21 volunteer students of the *Escuela de Ingeniería de Montes, Forestal y del Medio Natural* (Forestry and Natural Environment Engineering School) of the Technical University of Madrid. The experiment took place at the end of the semester, when it was assumed that all of the students had mastered the material. The experiment consisted in the identification of 20 additional woody species that had not been learned during the course. These species were randomly selected out of 40 remaining species in the herbarium. Although the set of selected species was new to the students at this time, it was assumed they had the competence skills needed to identify them.

The experiment consisted of five phases: (A) initially, the students were asked to take a diagnostic assessment (pre-test) to determine their previous knowledge of the 20 species. This assessment was performed in a manner similar to that of the final summative assessment, that is, students had to identify the species of a dried plant on a herbarium sheet; (B) the teacher explained the morphological characteristics of the new set of species in a laboratory class. The students could take notes and examine the plants; (C) at the end of the laboratory class, the students had additional time to reexamine the plants, revise their notes, or consult additional written material. In addition, half of the sheets (ten randomly selected) were labelled with a QR code. The code allowed the students to access (using an iPad) a formative assessment test that contained a *composed question* about that species (see Fig. 6). The composed questions were designed to guide students to ask the right questions to recognize the differential morphological features; (D) a diagnostic assessment (post-test) was

Fig. 6 Students taking a test in the laboratory using QR codes

repeated at the end, again using dried plant sheets of the same 20 species, so it can be assumed that the pre-test and post-test had the same level of difficulty. The sheets were different from those used in the previous phases to avoid cheating. Personal notes or additional material were not permitted; and (E) a questionnaire was used to determine the users' opinions. Table 3 summarizes the experiment.

The formative assessment consisted of ten *composed questions*, one for each species in the experimental group. Each question had four sub-questions. The structure of these questions was the same for the ten species. The first three questions addressed the differential morphological characteristics of the plant. The last sub-question asked for the scientific name. Each sub-question component also included elaborated feedback that was presented to the students according to their answer. Different feedback could be triggered according to each student's response. Feedback was usually presented in the case of a wrong answer, but it could also be presented if the answer was correct. Figure 7 shows an example of a composed question that was answered by a student. Notice that the student failed one of the three morphological questions, but correctly identified the species. Feedback was presented for sub-questions 2 and 3. The final score was not relevant to the experiment on the formative effect. The standard Cronbach's alpha was low (0.52), which means that the results of this test were not fully consistent. However, we were also interested in knowing whether the students identified a plant directly or by recognizing the morphological features. We tested this by splitting the formative test in two: the test that consisted of the 30 morphological questions was called T_m, and the test that consisted of the remaining ten identification questions was called T_i. The results were also compared to the post-test partial assessment of the ten species in the experimental group. Table 4 shows the average of the 21 students' aggregated scores.

However, the correlation coefficients (Table 5) show that there was a small correlation between the three assessments. The highest correlation was between the T_m and T_i scores, but none of the coefficients were significant ($p > 0.10$ in all cases). In our opinion, this indicates that the students did not apply deep morphological analysis to identify the plants. Further research in this line is needed.

The main hypothesis was that the plants that were studied with the support of the formative assessment were better recalled than those that were studied by only reviewing personal notes. Note that in this situation, the experimental group consisted of the set of ten species that have a QR code attached and that were included in the formative assessment and the control group consisted of the other ten species that do not have a QR code attached. The total number of identification cases can be obtained by multiplying the number of species and the number of students. Table 6 shows the results. The column labelled *#Ident* represents the total number of correct identifications in the pre-test and post-tests. The low percentage of identifications in the pre-test is not surprising because the 20 selected species were different from those that were previously studied. The column labelled *Score* represents the mean and standard deviations of the percentage of dried plants that were correctly identified. Both groups improved their results from the pre-test to the post-test. A two-tail

Table 3 Experimental design

Phase	Description	Duration (min)	#Students
A	Pre-test	20	22
B	Laboratory class	20	22
C	Self-study + formative assessment	30	21
D	Post-test	20	21
E	Questionnaire	5	20

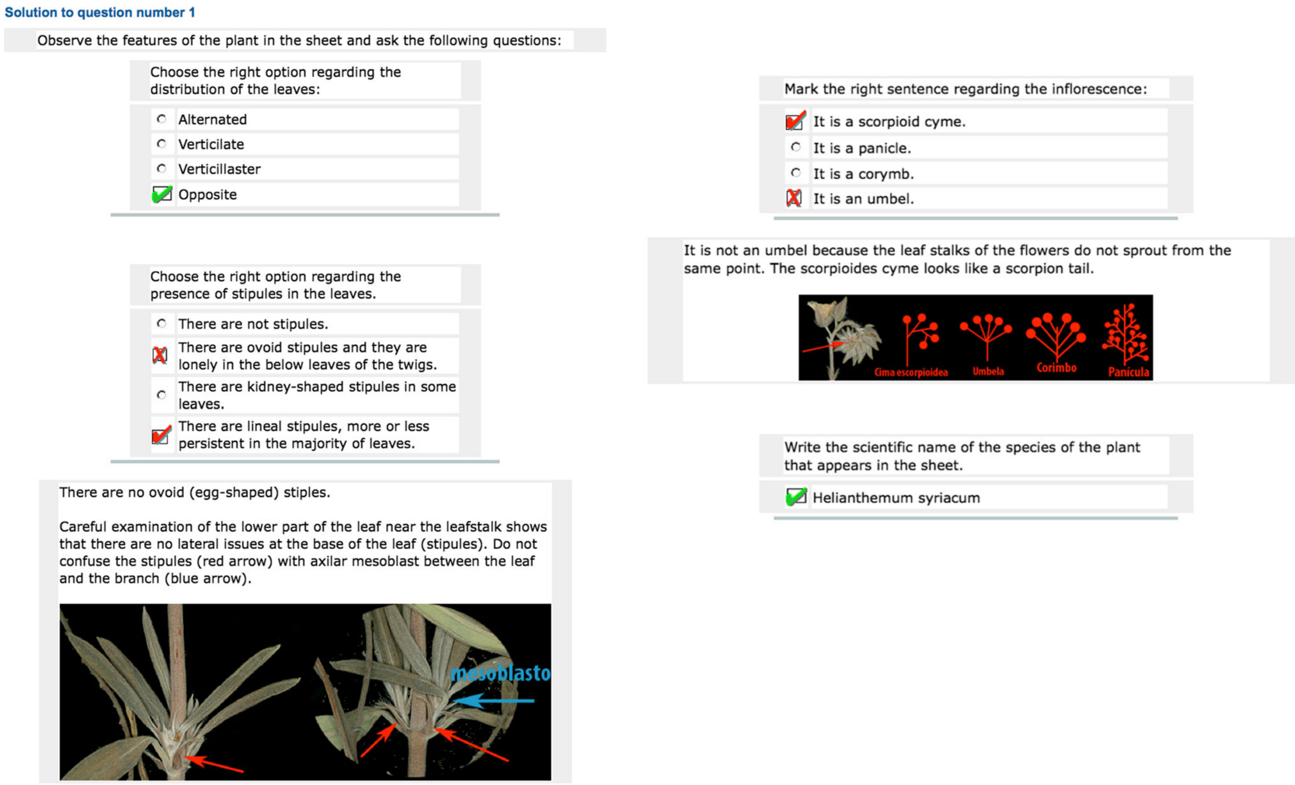


Fig. 7 Answered composed question with its four sub-questions. Correction and feedback is shown at the end of each sub-question

Table 4 Student scores in the three assessments

#Students	T_m score	T_i score	Post-test score (experimental group)
21	0.682 (0.209)	0.761 (0.139)	0.862 (0.186)

Mean (SD)

Table 5 Correlation coefficients between the three assessments

	T_m score	T_i score	Post-test score
T_m score	1	0.352	0.044
T_i score	0.352	1	0.306
Post-test score	0.044	0.306	1

paired t test was applied to compare the means of the learning gains of the experimental and control groups. The results clearly indicate that the experimental group outperformed the control group ($p = 1.27\mathrm{e}{-08}$), that is, the plants that were studied with the help of the formative assessment easier to learn than the others.

We evaluated these data to determine whether learning was related to the knowledge level of the students. We divided the student sample into two halves according to the scores in the post-test. We named one half L-students (those who performed below average in the post-test) and the other half H-students (those who performed above average). Table 7 shows the results. In this case, all the students improved their performance regarding the plants that were included in the experimental group. All the differences were highly significant (p value was always less than 0.0001). As expected, the L-students were those who

Table 6 Percentage of correct identifications (score)

	#Species	#Students	#Cases	Pre-test		Post-test	
				#Ident	Score	#Ident	Score
Experimental group of species	10	21	210	3	0.014 (0.119)	181	0.862 (0.346)
Control group of species	10	21	210	20	0.095 (0.293)	146	0.695 (0.461)

Mean (SD)

Table 7 Student skills and learning gains. Mean (SD)

	#Species	#Students	#Cases	Pre-test		Post-test	
				#Ident	Score	#Ident	Post-test score
H-students							
Experimental group	10	13	130	3	0.023 (0.151)	122	0.938 (0.241)
Control group	10	13	130	20	0.154 (0.362)	110	0.846 (0.362)
L-students							
Experimental group	10	8	80	0	0.000 (0.000)	59	0.738 (0.443)
Control group	10	8	80	0	0.000 (0.000)	36	0.450 (0.501)

improved more in terms of absolute difference between the experimental and control group because they had greater room for improvement.

Finally, 20 students completed an anonymous questionnaire. These students were different from the students who answered the questionnaire for the fresh plant experiment last year (see Table 2). All the questions were answered using a 1–5 Likert scale. Depending on the question, 1 means "none", "very low", or "completely disagree", and 5 means "all", "very much", or "completely agree". The results are shown in Table 8 with the mean and standard deviations. The results are consistent with those that were obtained the previous year. The overall evaluation (Question-18) of the system and the activity was high, with an average of 3.90, which was slightly lower than the previous year in the scenario using fresh plants. There was a positive correlation between answers to Question-18 and Question-4 ($\rho = 0.61$) and Question-5 ($\rho = 0.71$). No other significant correlation was found between Question-18 and any other question, including Question-1 or Question-2, that is, no relationship was found between the overall evaluation and previous experience with the iPad or smartphones.

Almost all the students indicated that they preferred the assessment based on fresh plants (Question-14, 4.74) followed by assessment with dried plants (Question-15, 4.53). Consistent with this preference, they concluded that assessment with dried plants was appropriate for formative assessments (Question-16, 4.05) and for summative assessments (Question-17, 3.65). If we compare these

Table 8 Results of the anonymous questionnaire after the formative test with dried plants

Question	1	2	3	4	5	Mean (SD)
About the student						
1. My experience of using an iPad is…	4	2	5	5	4	3.15 (1.39)
2. My experience with mobile devices is…	1	0	5	9	5	3.85 (0.96)
3. My experience with computer-based evaluation systems is …	0	4	11	5	0	3.05 (0.67)
4. I think that the use of information technologies for learning is helpful	0	0	6	12	2	3.80 (0.60)
About the activity						
5. I have enjoyed taking the location-aware test	0	1	4	11	4	3.90 (0.77)
6. It required extra effort to do it	1	3	9	7	0	3.10 (0.83)
7. I enjoyed this activity	0	0	8	10	2	3.70 (0.64)
8. The test at the end of the class helped me to learn to identify the plants	0	0	3	7	10	4.35 (0.73)
About the system						
9. The system worked fine. I was able to complete the test and the system always behaved as expected	1	7	5	4	3	3.05 (1.16)
10. The system interface is intuitive. I always found the buttons and options whenever I needed them	0	2	6	5	7	3.85 (1.01)
11. At any time I knew where I was and what was left to finish the test.	0	3	3	8	6	3.85 (1.01)
12. The system is easy to use	0	2	1	8	8	4.16 (0.93)
13. I prefer to take the test with the iPad and dry plants than with photographs in the computer	0	1	4	5	9	4.16 (0.93)
14. I prefer to take the test with the iPad and fresh plants than with photographs in the computer	0	0	1	3	15	4.74 (0.55)
15. I prefer to take the test with fresh plants than with dried plants	0	1	1	4	13	4.53 (0.82)
16. This type of test is appropriate for practicing and learning	0	0	3	13	4	4.05 (0.59)
17. This type of test is appropriate for final assessment	0	1	6	12	1	3.65 (0.65)
About the system						
18. Overall evaluation about the use of the system and the activity	0	0	5	12	3	3.90 (0.62)

results with those that were obtained last year with fresh plants, it can be concluded that in both cases the students intuitively indicated that these types of assessments are better suited to formative assessment than to summative assessments and that they preferred fresh plant identification to dried plant identification, and dried plant identification to photograph identification. An interesting issue concerns the perception of learning gains (Question-8, 4.35). Most of the students strongly agreed that the test at the end of the laboratory class helped them *very much* to learn to identify the plants.

Identification of Fresh Plants: Measuring the Effect of Technology-Enhanced Assessment

The aim of this experiment was to examine the benefits of using this technology in the learning process. The specific research questions were: *What new educational opportunities does this technology provide? Does this technology provide an educational benefit above what is provided by existing teaching tools?* A parallel objective was to determine the students' subjective perceptions of the effect of the technology and to collect their opinions about the usability of the system.

The experiment included 36 volunteer students of the *Escuela de Ingeniería de Montes, Forestal y del Medio Natural* (Forestry and Natural Environment Engineering School) of the Technical University of Madrid. The experiment took place at the end of the semester, when it was assumed that all the students had mastered the material. The experiment consisted of the identification of 24 additional woody species that were not learned during the course. At this time, the students were assumed to have the competence skills that were needed to identify the species.

The experiment consisted of five phases: (A) initially, the students were asked to take a diagnostic assessment (pre-test) to determine their previous knowledge of the 24 species randomly selected from the 48 available at the arboretum. This assessment was performed in a manner similar to that of the final summative assessment, that is, students had to identify the species of a dried plant on a herbarium sheet; (B) the students went to the arboretum to study the 24 species. The test was split into two halves that were taken one after the other; the order was also changed from one student to another according to iPad availability. The first half of the test contained half of the specimens (12 randomly selected specimens) that were labelled with a number. Students had to complete a paper-and-pencil test, using their notebook. The notebook contains a page for each species including its description and explanatory drawings (see Fig. 8). The other half of the species (also randomly selected) was labelled with a QR code. Using an iPad, the code allowed access to a formative assessment test that contained a *composed question* about the species (see Fig. 8). The composed questions were designed to guide students towards asking the right questions to recognize the differential morphological features. The students received delayed and elaborated feedback from a teacher in the case of plants with an attached number, and automatic immediate elaborated feedback in the case of the species labelled with a QR code. Note that, in the former case, unless a teacher accompanied every student it would have been impossible to accomplish immediate feedback without promoting cheating; (C) a diagnostic assessment (post-test) was repeated at the end, again using dried plant sheets of the same 24 species. The sheets were different from those that were used in the previous phases to avoid cheating. Due to its parallel design, it can be assumed pre-test and post-test had the same level of difficulty and thus the scores are comparable. Personal notes or additional material were not permitted in the pre-test and post-test; (D) a questionnaire was given to determine the users' opinions. Table 9 summarizes the results of the experiment.

The experiment took over 2 weeks to complete because no more than three students took the experiment at the same time. This restriction was imposed for two reasons: (1) only three iPads were available; (2) student communication in the arboretum had to be restricted to avoid cheating and thus avoid the introduction of corrupt data. To guarantee equal conditions across time for all students, the QR codes and numbers in the arboretum were regularly changed from one specimen to another of the same species.

The main hypothesis was that the plants that were studied with the support of technology are better recalled than those that studied with traditional methods. Note that in this situation, the experimental group consisted of the set of 12 species that were labelled with the QR code, and the control group consisted of the set of 12 species that were labelled with a number. The total number of identification cases can be obtained by multiplying the number of species and the number of students. In the experimental group, this number was less than expected because some students had problems with the Wi-Fi connection. Table 10 shows the results. The column labelled *#Ident* represents the total number of correct identifications in the pre-test and post-tests. The column labelled *Score* represents the mean and standard deviations of the percentage of dried plants that were correctly identified. The percentage of identifications in the pre-test was almost the same in the experimental and control group, which indicates that the random selection of species was done correctly. Both groups improved their results from pre-test to post-test. A two-tail paired t test was applied to compare the means of the learning gains (the difference in the number of plants correctly identified) in the experimental and control groups. The results clearly

Ericaceae

Erica australis L.
Nombre vulgar: brezo, brezo negro, brezo colorado.

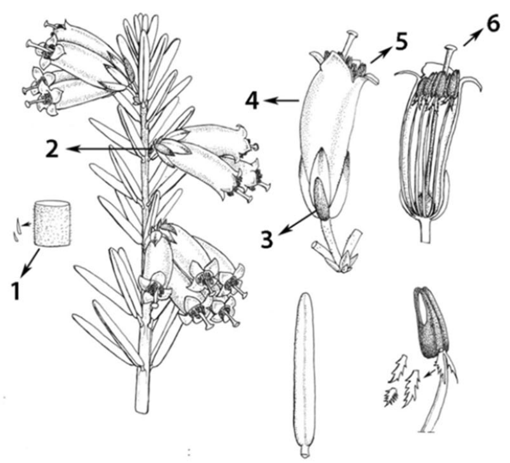
Mata perennifolia de talla media o alta. Ramillas terminales finamente tomentosas. Hojas verticiladas; lineares; muy revolutas. Inflorescencias distribuidas por todo el conjunto de la mata, con aspecto de umbelas de radios imperceptibles porque están tapados por una brácteas purpúreas. Flores con la corola rosada o rojiza; tubular; larga (6 – 8,5 mm) y de cuyo extremo sobresalen algo las puntas de las anteras. Anteras inclusas. Estilo exerto con un estigma claviforme. El fruto es una pequeña cápsula que permanece dentro de la corola ya seca.

Principales caracteres diferenciales *de visu*:

- Ramillas terminales con un tomento poco visible, muy corto (1).
- Hojas sin características relevantes.
- Flores en grupos reducidos, sin pedúnculos visibles (2).
- Brácteas sepaloideas, largas (3) y rosadas.
- Corolas tubulares (4), rosada o rojiza en la antesis.
- Estambres inclusos (5).
- Estigmas claviformes (6).

Altitud (m)	Litología	Dominios de vegetación							
		AM	T	C	S	E	H	IZ no HD	IZ HD
0 – 2000	calcífuga			+	+	+			

Se encuentra en diversos tipos de brezales y formando parte de un conjunto amplio de formaciones arbóreas, como encinares, alcornocales, madroñales, rebollares de *Quercus pyrenaica*, hayedos, pinares de *Pinus pinaster*.



Cistaceae

Cistus populifolius L.
Nombre vulgar: jara cervuna, estepa.

Mata perennifolia, de talla alta aunque ocasionalmente puede alcanzar hasta algo más de 2 m. Ramillas gruesas sin la corteza desprendiéndose en tiras. Ramillas terminales de hirsutas a glabrescentes. Hojas opuestas; sin estípulas; con un peciolo largo y un limbo ancho (3 – 5 cm), cordado y agudo, en el que se puede apreciar un retículo de nervios característico y con el margen liso o débilmente ondulado. Inflorescencia cimosa y larga. Flores con 5 sépalos y 5 pétalos blancos y grandes (1,5 – 3 cm). El fruto es una cápsula glabrescente, con 5 valvas loculicidas.

Principales caracteres diferenciales *de visu*:

- Hojas opuestas, largamente pecioladas, con el limbo ancho y de nerviación reticulada (1).
- Sépalos acorazonados (2).
- Pétalos blancos.
- Cápsula con 5 valvas glabrescentes (3).

Altitud (m)	Litología	Dominios de vegetación							
		AM	T	C	S	E	H	IZ no HD	IZ HD
100 – 1000 (1500)	calcífuga				(+)	+			

Se encuentra formando parte de matorrales mixtos como las manchas, rara vez monoespecíficos (poca superficie) y formado parte de pinares de *Pinus pinaster*, *Pinus pinea*, *Pinus halepensis*, encinares y alcornocales.

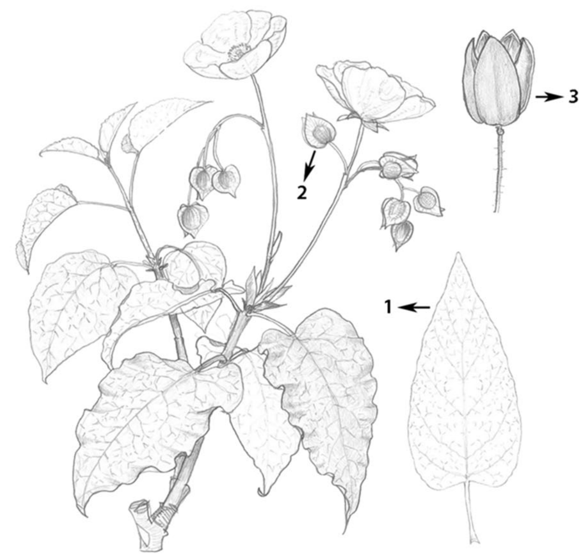


Fig. 8 Students' notebook

Table 9 Experimental design

Phase	Description	Duration (min)	#Students
A	Pre-test	20	36
C	Field formative assessment	60	36
D	Post-test	20	36
E	Questionnaire	5	36

indicate that the experimental group outperformed the control group ($p < 0.0001$), that is, that the plants that were studied with the help of technology were easier to recall than the others.

We evaluated these data to determine whether learning was related to the knowledge level of the students. We divided the student sample into two halves according to the score in the post-test. We named one half L-students (those who performed below average in the post-test) and the other half H-students (those who performed above average). Table 11 shows the results. In this case, all the students improved their performance regarding the plants that were included in the experimental group. All the differences were significant (p value was always less than 0.05). As expected, the L-students were those who improved more in terms of absolute difference between the experimental and control group because they had greater room for improvement.

A complementary result was obtained from the formative assessment itself. In this experiment, the education instrument was also a test divided in two parts. One half (the control group) was completed with paper and pencil and the help of a notebook, whereas the other half (the experimental group) was completed without these aids. Both parts contained three questions on the morphological aspects of every species, and a fourth question addressed species identification. In the case of the paper-and-pencil test, the answers to the morphological questions were meaningless, because if the students had identified the

Table 10 Percentage of correct identifications (score)

	#Species	#Students	#Cases	Pre-test		Post-test	
				#Ident	Score	#Ident	Score
Experimental group of species	12	36	418	253	0.605 (0.486)	393	0.940 (0.237)
Control group of species	12	36	432	266	0.616 (0.489)	341	0.789 (0.408)

Mean (SD)

Table 11 Student skills and learning gains

	#Species	#Students	#Cases	Pre-test		Post-test	
				#Ident	Score	#Ident	Post-test score
H-students							
Experimental group	12	20	480	190	0.819 (0.385)	227	0.978 (0.145)
Control group	12	20	472	182	0.758 (0.428)	205	0.854 (0.353)
L-students							
Experimental group	12	16	80	63	0.339 (0.473)	166	0.892 (0.310)
Control group	12	16	80	84	0.438 (0.496)	136	0.708 (0.455)

Mean (SD)

species, then they were able to complete the information from the notebook. However, in the electronic version, they had to answer each question independently. For this reason, we only focussed on the correct identification of the species. Table 12 shows the results.

These results indicate that the recognition was far easier for the students with the notebook. Although they could browse the notebook and select the most similar drawing, they did not pay sufficient attention to the morphological details that are important for identification. With the technology-enhanced formative assessment, they were forced to concentrate on details and feedback was presented immediately, which is known to be the best strategy.

Finally, 24 students completed an anonymous questionnaire. These students were different from the students who answered the questionnaire for the fresh plant experiment last year (see Tables 2, 8). All the questions were answered using a 1–5 Likert scale. Depending on the question, 1 means "none", "very low," or "completely disagree", and 5 means "all", "very much", or "completely agree". The results are shown in Table 13 with the mean and standard deviations. The results are consistent with those that were obtained the previous year. The overall evaluation (Question-18) of the system and the activity was high, with an average of 4.25, which was even higher than the previous year in the scenario using fresh plants. There was a positive correlation between answers to Question-18 and Question-8 ($\rho = 0.71$) and Question-14 ($\rho = 0.77$). No correlation was found between Question-18 and Questions-1 or 2, that is, no relationship was found between the overall evaluation and previous experience with the iPad or smartphones.

Almost all the students indicated that they prefer the assessment based on fresh plants, rather than with dried plants (Question-15, 4.42). Consistent with this preference, they concluded that the assessment with fresh plants was appropriate for formative assessments (Question-16, 4.63) and even for summative assessments (Question-17, 4.25). These results are consistent with those from Experiments 1 and 2. An interesting issue concerns the perception of learning gains (Question-8, 4.35). Most of the students strongly agreed that the test at the end of the laboratory class helped them *very much* to learn to identify the plants;

Table 12 Percentage of correct identifications during the activity (score)

	#Species	#Students	#Cases	Formative assessment	
				#Ident	Score
Experimental group of species	12	36	418	208	0.498 (0.500)
Control group of species	12	36	432	347	0.803 (0.398)

Mean (SD)

Table 13 Results of the anonymous questionnaire after the formative test with fresh plants

Question	1	2	3	4	5	Mean (SD)
About the student						
1. My experience of using an iPad is…	1	3	8	4	6	3.58 (1.18)
2. My experience with mobile devices is…	1	1	7	7	6	3.74 (1.05)
3. My experience with computer-based evaluation systems is …	0	4	9	6	4	3.46 (0.98)
4. I think that the use of information technologies for learning is helpful	0	1	5	9	7	3.96 (0.81)
About the activity						
5. I enjoyed taking the paper-and-pencil test in the arboretum	0	1	9	8	4	3.71 (0.81)
6. I enjoyed taking the iPad test in the arboretum	1	1	2	14	3	3.88 (0.90)
7. I learned by taking the paper-and-pencil test in the arboretum	2	2	6	8	4	3.42 (1.14)
8. I learned by taking the iPad test in the arboretum	0	0	2	8	10	4.46 (0.66)
About the system						
9. The system worked fine. I was able to complete the test and the system always behaved as expected	1	4	4	7	6	3.67 (1.20)
10. The system interface is intuitive. I always found the buttons and options whenever I needed them	0	0	3	12	6	4.17 (0.64)
11. At any time I knew where I was and what was left to finish the test	1	2	7	7	4	3.63 (1.06)
12. The system is easy to use	0	0	1	13	6	4.25 (0.53)
13. I prefer to take the test with the iPad rather than with paper and pencil in the arboretum	0	1	4	5	9	3.54 (1.14)
14. The feedback from the iPad test was very useful for learning	3	0	6	10	3	4.29 (0.86)
15. I prefer to take the test with fresh plants than with dried plants	0	0	3	5	12	4.42 (0.88)
16. This type of test is appropriate for practicing and learning	0	0	1	4	15	4.63 (0.58)
17. This type of test is appropriate for final assessment	0	1	3	7	10	4.25 (0.85)
About the system						
18. Overall evaluation about the use of the system and the activity	0	0	3	11	6	4.25 (0.68)

Table 14 Summary results of questionnaires

Question	Experiment 1 2012 (fresh plants)	Experiment 2 2013 (dry plants)	Experiment 3 2014 (fresh plants)
I think that the use of information technologies for learning is helpful	4.00 (0.38)	3.80 (0.60)	3.96 (0.81)
I prefer to take the test with dry plants than with photographs.		4.16 (0.93)	
I prefer to take the test with fresh plants than with photographs.	4.77 (0.53)	4.74 (0.55)	
I prefer to take the test with fresh plants than with dry plants		4.53 (0.82)	4.42 (0.88)
This type of test is appropriate for practicing and learning	4.73 (0.63)	4.05 (0.59)	4.63 (0.58)
This type of test is appropriate for final assessment	4.50 (0.74)	3.65 (0.65)	4.25 (0.85)
Overall evaluation about the use of the system and the activity	4.36 (0.58)	3.90 (0.62)	4.25 (0.68)

Mean (SD)

this was significantly different to the paper-and-pencil assessment. Overall, the students were in agreement that the system worked well and that it was easy to use. They highlighted immediate feedback as the most significant feature. On the other hand, some students complained about the Wi-Fi connection, which might be an explanation for a relatively low value that some of them assigned to Question-9.

Table 14 summarizes the results obtained from the three questionnaires after each experiment. Although the population differed from 1 year to another, the responses are consistent.

Conclusions

Continuous formative assessment is an effective way to acquire the skills required for plant identification. Based on plant photographs, this method has been used since 2000 in the botany course at the *Escuela de Ingeniería de Montes,* *Forestal y del Medio Natural* (Forestry and Natural Environment Engineering School) of the Technical University of Madrid. Using immediate elaborated feedback, including text and photographs, formative assessment produces measurable gains in test performance. However, assessment conditions are different from the identification skills that are required for the final summative assessment of the botany course, which is based on the identification of dried plants, and also different from real-world practitioner work, which is based on the identification of fresh plants in their natural environments. To reduce this gap, we developed an extension of the Siette assessment framework that uses geographical positioning and/or QR codes to trigger questions attached to a specific location.

Our conclusion is that the system is technologically robust. All the students were able to take and complete the tests in a reasonable amount of time. From the point of view of usability and usefulness, the opinions that were anonymously expressed by the students were also very favourable. We consider that the application presented in this article considerably broadens the applicability of automatic assessment techniques. With this technique, testing can be performed in scenarios that are more similar to those found in real-world situations.

This technology also provides new educational opportunities. The traditional way to learn how to identify dried or fresh plants is to use a field notebook and to practice in a laboratory with a herbarium or an arboretum with fresh plants. A review of the literature finds non-technological proposals based on practice with dichotomous keys, and some technological proposals that help to recognize plants but do not ensure that the student has made a correct identification; this can only be ensured by a teacher accompanying a group of students and proposing questions that lead to correct identification. These questions include: "*What is the disposition of the leaves?", "Are there stipples?*" or "*What is the type of inflorescence?*". The system described allows the implementation of this Socratic teaching style. There are two advantages: (1) the students can practice on their own, without the need for a teacher to pose questions and correct answers; and (2) the instruction is personalized for each student according to his/her answers.

Three experiments were designed to evaluate the system, measure system performance, and compare the learning gains obtained with other traditional practicing methods. The effect of formative assessment was evaluated by identifying dried and fresh plants. In all cases, the main conclusions were that the students learn by using the system. Taking into account that the aim of the system is not to replace the teacher but to allow the students further practice, the control groups were designed using practice methods that do not require the presence of the teacher. According to previous results in cognitive psychology, we have shown that taking an automatic formative assessment after a laboratory class qualitatively increases retention compared to approaches based on simply revising notes. Our final experiment also shows that students who take on-site formative assessment with immediate feedback obtain better results than those obtained with formative assessment using traditional methods. Furthermore, this technology-enhanced system was highly appreciated by students and their attitudes towards the system were very positive.

Summing up, we have found that technology can enhance the practice of plant identification and drawn the following conclusions: (1) Almost all previous studies in this field (Ohkawa 2000, Uno 2009, Silva et al. 2011, Stagg and Donkin 2013) have emphasized the importance of outdoor education with fresh plants. We have found that it is easier for students to identify living plants than dried plants or photographs. This finding was supported by objective data and by the students' subjective impressions. The students' attitude towards the system was positive, and their opinions indicate that they will accept the system for formative and even summative assessment; (2) practicing plant identification with dried or fresh plants improved the results obtained in the post-test based on dried plant identification (the post-test has similar conditions to the actual final exam). Technology can help to improve the way that students practice plant recognition. It is already known that testing after studying has a beneficial effect on learning (McDaniel et al. 2007; Karpicke and Roediger 2008) and that immediate feedback produces the best results (Shute 2008). Using the system presented in this paper, students can take advantage of this effect. Experiments 2 and 3 showed that the students improved more by using the system than using alternative classic methods. Without technology, it would be almost impossible to implement testing with immediate feedback; and (3) most of the students follow a holistic approach for plant identification. This aspect has been recently addressed by Kirchoff et al. (2014) and was also supported by the results of Experiments 2 and 3, which are included in this study. However, a variety of procedures related to learning objectives should be taken into account. For example, due to the major role played by morphological features in plant identification, it would be of interest to explore procedures that could guide the recognition of these features (such as the immediate feedback provided by our system). These procedures could make the results of learning more robust and thus lead to better long-term results. However, this issue is beyond the scope of this work and will require future research.

On the other hand, the use of technology may have some limitations: (1) it requires a considerable effort by the teacher to prepare the questions and feedback, and to print and distribute the QR-code labels. Of course, the number of students is the key factor in adopting this technology. If the course is repeated then all the material can be reused; (2) there is a significant financial burden on either the students or the institution, which might reduce educational opportunities for some students. However, having three iPads to share was sufficient for the experiments. As the teacher does not need to be present, the students do not need to practice at the same time; (3) the addition of an iPad or smartphone might represent a distraction from the subject matter at hand, and interfere with hands-on interaction with the plants. Although this situation is possible, we did not observe it in our experiments. The system is set up for self-regulated learning, so this factor depends on the maturity of the students. The experimental results and response to the questionnaires suggest that this would not be a major concern at higher educational levels.

This article forms part of on going research. The positive results of these pilot studies encourage us to implement a large-scale framework for self-regulated formative assessment. Our plan is to label all of the sheets in the herbarium and all of the plants in the arboretum with a QR code that will be linked to a set of questions about that plant. Students will be allowed to access these resources freely and will receive immediate feedback. We are also considering the possible use of a context-aware system for summative assessment. Currently, the final assessment of the botany course consists of three parts: an essay on a selected topic of the curriculum; a test of general concepts and plant photograph identification, which has already been taken using a computer-based system; and randomly selected plant identification based on dried plants. The automatic assessment of dried and fresh plant identification is unlikely to become part of a formal final assessment at this time because it is difficult to guarantee the controlled conditions that are needed to avoid cheating in an open environment. At this time, another limiting factor is the availability of mobile devices for students, but we hope that this will change in the near future.

References

Anderson LW, Krathwohl DR, Bloom BS (2001) A taxonomy for learning, teaching, and assessing a revision of Bloom's taxonomy of educational objectives. Longman, New York

Biggs JB (2011) Teaching for quality learning at university: what the student does. McGraw-Hill Education, UK

Black P, William D (2009) Developing the theory of formative assessment. Educ Assess Eval Account 21(1):5–31

Bloom BS, Krathwohl DR, Masia BB (1956) Taxonomy of educational objectives: the classification of educational goals. D. McKay, New York

Buchanan T (2000) The efficacy of a World-Wide Web mediated formative assessment. J Comput Assist Learn 16(3):193–200

Conejo R, Guzmán E, Millán E, Trella M, Pérez-De-La-Cruz JL, Ríos A (2004) SIETTE: a web-based tool for adaptive testing. Int J Artif Intell Educ 14(1):29–61

Cooper B, Cowie B (2010) Collaborative research for assessment for learning. Teach Teach Educ 26(4):979–986

Crowe A, Dirks C, Wenderoth MP (2008) Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. CBE-Life Sci Educ 7(4):368–381

Dufresne R, Gerase W (2004) Assessing-to-learn: formative assessment in physics instruction. Phys Teach 42:428–433

Fabienne M, van der Kleij M, Eggen TJHM, Timmers CF, Veldkamp BP (2012) Effects of feedback in a computer-based assessment for learning. Comput Educ 58(1):263–272

Gardner J, Belland BR (2012) A conceptual framework for organizing active learning experiences in biology instruction. J Sci Educ Technol 21(4):465–475

Gikandi JW, Morrow D, Davis NE (2011) Online formative assessment in higher education: a review of the literature. Comput Educ 57(4):2333–2351

Guzman E, Conejo R, Perez de la Cruz JL (2007) Improving student performance using self-assessment tests. IEEE Intell Syst 22:46–52

Huang H-M (2002) Toward constructivism for adult learners in online environments. Br J Educ Technol 33:27–37

Huang HW, Wu CW, Chen NS (2012) The effectiveness of using procedural scaffoldings in a paper-plus-smartphone collaborative learning context. Comput Educ 59:250–259

Hwang GJ, Chang HF (2011) A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. Comput Educ 56(4):1023–1031

ISO (2006) Information technology—automatic identification and data capture techniques—QR Code 2005 bar code symbology specification. Technical Report ISO/IEC 18004:2006 International Organization for Standardization

Jordan S (2012) Student engagement with assessment and feedback: some lessons from short-answer free-text e-assessment questions. Comput Educ 58(2):818–834

Karpicke JD, Roediger HL III (2008) The critical importance of retrieval for learning. Science 319:966–968

Kato H, Tan KT (2007) Pervasive 2D barcodes for camera phone applications. IEEE Pervasive Comput Comput Mob Ubiquitous Syst 6:76–85

Kirchoff BK, Delaney PF, Horton M, Dellinger-Johnston R (2014) Optimizing learning of scientific category knowledge in the classroom: the case of plant identification. CBE-Life Sci Educ 13(3):425–436

Klionsky DJ (2008) The quiz factor. CBE-Life Sci Educ 7(3):265–266

Laine TH, Vinni M, Sedano CI, Joy M (2010) On designing a pervasive mobile learning platform. Res Learn Technol 18(1):3–17

Law C, So S (2010) QR codes in education. J Educ Technol Dev Exchange 3:85–100

Lee SW-Y, Tsai C-C (2013) Technology-supported learning in secondary and undergraduate biological education: observations from literature review. J Sci Educ Technol 22:226–233

Lucke U, Rensing C (2014) A survey on pervasive education. Pervasive Mob Comput 14:3–16

McDaniel MA, Anderson JL, Derbish MH, Morisette N (2007) Testing the testing effect in the classroom. Eur J Cogn Psychol 19:494–513

Momsen J, 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 Sci Educ 12(2):239–249

Ohkawa C (2000) Development of teaching materials for field identification of plants & analysis of their effectiveness in science education. Am Biol Teach 62(2):113–123

Ozcelik E, Acarturk C (2011) Reducing the spatial distance between printed and online information sources by means of mobile technology enhances learning: using 2D barcodes. Comput Educ 57:2077–2085

Palmer EJ, Devitt PG (2007) Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Res Pap BMC Med Educ 7(1):49–56

Preszler RW, Dawe A, Shuster CB, 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 Sci Educ 6(1):29–41

Randler C (2008) Teaching species identification—a prerequisite for learning biodiversity and understanding ecology. Eurasia J Math Sci Technol Educ 4:223–231

Romero C, Ventura S, De Bra P (2009) Using mobile and web-based computerized tests to evaluate university students. Comput Appl Eng Educ 17(4):435–447

Santos P, Hernández-Leo D, Blat J (2014) To be or not to be in situ outdoors, and other implications for design and implementation, in geolocated mobile learning. Pervasive Mob Comput 14:17–30

Scott GW, Goulder R, Wheeler P, Scott LJ, Tobin ML, Marsham S (2012) The value of fieldwork in life and environmental sciences in the context of higher education: a case study in learning about biodiversity. J Sci Educ Technol 21(1):11–21

Shute VJ (2008) Focus on formative feedback. Rev Educ Res 78(1):153–189

Silva H, Pinho R, Lopes L, Nogueira AJ, Silveira P (2011) Illustrated plant identification keys: an interactive tool to learn botany. Comput Educ 56(4):969–973

Stagg BC, Donkin M (2013) Teaching botanical identification to adults: experiences of the UK participatory science project 'Open Air Laboratories'. J Biol Educ (ahead-of-print) 1–7

Stanger-Hall KF (2012) Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. CBE-Life Sci Educ 11(3):294–306

Tractenberg RE, Gushta MM, Mulroney SE, Weissinger PA (2013) Multiple choice questions can be designed or revised to challenge learners' critical thinking. Adv Health Sci Educ 18(5):945–961

Uluyol C, Agca RK (2012) Integrating mobile multimedia into textbooks: 2D barcodes. Comput Educ 59:1192–1198

Uno GE (2009) Botanical literacy: What and how should students learn about plants? Am J Bot 96(10):1753–1759

Zoller U (2001) Alternative assessment as (critical) means of facilitating HOCS-promoting teaching and learning in chemistry education. Chem Educ Res Pract Eur 2(1):9–17

Zoller U, Tsaparlis G (1997) Higher and lower-order cognitive skills: the case of chemistry. Res Sci Educ 27(1):117–130