

Chapter 7

Assessing Student Learning Online

Overcoming Reliability Issues

Stephen D. Arnold

1 Introduction

Assessment is the process of collecting, synthesizing, and interpreting information in order to make a decision (Airasian and Russell 2007). In the college class, this decision inevitably not only includes students' grades but it also ties to teachers' success in covering the content effectively, preparedness of students, and the accreditation worthiness of programs in academically preparing students for the larger picture. Value in any instructional system comes from assessment; what is assessed in a course or a program is generally associated with value; what is valued becomes the focus of activity (Swan et al. 2007). Effective assessment typically includes ongoing "formative assessment" checkpoints and end-of-term "summative assessment." Instructors signal what knowledge skills and behaviors they believe are most important by assessing them, while students quickly respond by focusing their learning accordingly (Swan et al. 2007). The end-of-course assessment method, and more specifically the requirements that underlie this assessment mode, make a difference to the outcome (Struyven et al. 2006). Considering that stability reliability equates to consistency of test results over time (Popham 2011), it is safe to assume that the processes involved in assessment from day 1 to the end of the term can have a vast impact on reliability. Systematic processes of reliable assessment do not end after a test is developed, especially in the online sector.

In online, asynchronous courses, whereby the students and instructor do not meet, obtaining reliable assessment measures becomes more difficult than in a traditional face-to-face (F2F) class. It is important to collect several pieces of information about the performance being assessed to increase reliability (Airasian and Russell 2007). Although it is possible as an instructor to elicit online quizzes, papers, and projects from students, there is still the dilemma of determining who is (and how many are) involved in the submission of the common assessment items.

S. D. Arnold (✉)

University of Arizona South, UA Science & Technology Park, 9040 South
Rita Rd., Tucson, AZ 85747, USA
e-mail: sdarnold@email.arizona.edu

Today's movement toward exponentially higher online enrollments and the ensuing assessment issues is best illustrated by the Massive Open Online Course (MOOC) euphoria. Aside from the extremely low completion rates, MOOCs are faced with the challenge of how to effectively assess thousands of students enrolled in these courses each semester. Reliability of assessment in the MOOC scenario is not so critical with regard to grading since the vast majority of students are taking the courses in the noncredit, open learning capacity. With the talk of moving toward degree fulfilling, credit-earning MOOCs, reliability is a major hurdle that will have to be addressed, however. The viewpoint of one instructor from a highly ranked school may capture the magnitude of reliability issues faced in the move toward larger online enrollments when he said that he likes the idea of drilling students with online quizzes, but his own MIT students would have to work on theirs in a classroom with a proctor (Kolowich 2013a).

There are issues with the course management system (CMS) interfaces that influence testing processes in a manner that impacts results. Clearly, different approaches will have to be implemented when teaching a class of 20 local students versus a class enrollment of 33,000 globally, which is the MOOC mean according to Kolowich (2013a). As a means to strengthen assessment reliability and foster students' creative engagement, the use of alternative digital pontifications must be examined and discussed as a viable means to foster more reliable assessment outcomes for students and instructors in lower-enrollment, local online courses. In the higher-enrollment courses, in-person or virtually proctored assessments must be explored.

2 Input and Output

It is possible to look at the process of developing one's content knowledge as "input" and demonstrating what one knows as "output." More commonly, this is referred to as learning and assessment. Hunter (2004) equated the terms with input of information into the students' cognitive learning processes and output of information in a mastery of the learning-objective sense so that proper assessment may occur. Output is also associated with the active process of learning, whereby the process of output draws heavily upon the content knowledge students experienced through the input process reinforcing the learning (Arnold and Moshchenko 2009). When students are given the opportunity to produce a tangible product or demonstrate something to an audience, their willingness to put forth quality increases (McTighe 1996). As preservice teachers in technology for educators courses, students' input comes through the convergence of four primary areas: (1) K-12 subject matter (math, science, etc.), (2) pedagogical knowledge (how to teach effectively with technology), (3) technological knowledge (usually through extensive technology tutorials), and (4) educational technology foundations (content knowledge that pertains to why we use technology). Students synthesize this information and produce digital and tangible output which further reinforces the input process while demonstrating learning growth (see Fig. 7.1).

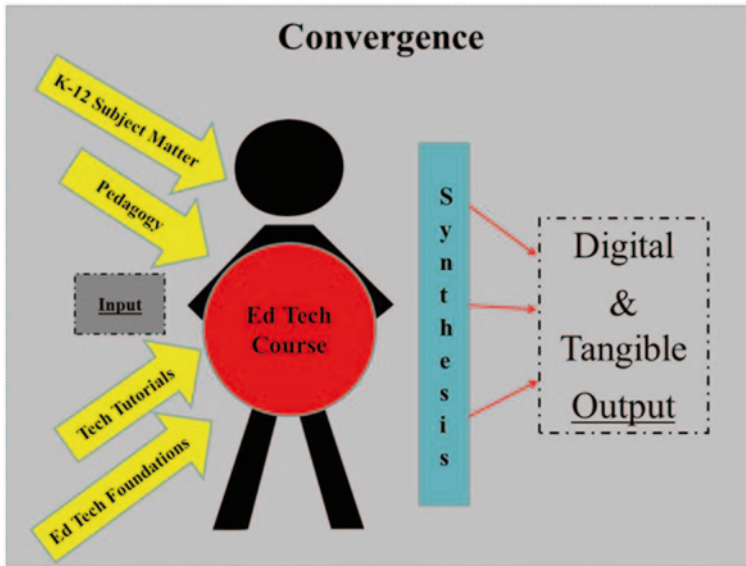


Fig. 7.1 Convergence of multiple knowledge tracks maximizes the input/output process

3 Assessment

In any course, measuring the growth students have made toward the objectives is critical to determine the effectiveness of instruction. As instructors, we have to be certain that our efforts are resulting in optimum outcomes for students. In higher education, written exams are often the assessment means of choice due to large numbers of students and limited instructor time. Depending upon the academic major, there is oftentimes professional dissonance between the weight of project, presentation, discussion, prose, and exam-evidenced proficiency being required of students. Final course grades and exams are the most common measures of learning outcomes for seniors across majors, however (NSSE 2010). Scheduled test events tend to increase students' study-time efficiency (McKenzie 1979).

In addition to measuring the level of proficiency growth, the assessment process further stimulates students' repetition and engagement with the course content. Highly familiar, meaningful stimuli subjected to increased processing time are directly correlated to increased retention of the stimuli (Craik and Lockhart 1972). More frequent assessment episodes, such as weekly quizzes, provide an increase in focused processing time. Roediger et al. (2011) identified ten benefits of testing:

1. Retrieval induced by testing facilitates later retention
2. Identifies gaps in knowledge
3. Causes students to learn more from the next learning episode
4. Produces better organization of knowledge
5. Improves transfer of knowledge to new contexts

6. Facilitates retrieval of information that was not tested
7. Improves metacognitive monitoring
8. Prevents interference from prior material when learning new material
9. Provides feedback to instructors
10. Encourages students to study

Assessment is an important opportunity for student learning as well as a means for instructors to judge student performance and assign grades (Thorpe 1998).

3.1 *Types of Assessment*

There are many categorizations of assessment. First, it is possible to make a distinction between assessment and test, with the first being the process of determining the learning gains, and the second being the instrument or measurement tool for gathering the data. Some use the terms evaluation and assessment in a similar manner. It is important to consider the time-based snapshots of learning, which are addressed by ongoing incremental *formative* measurements and the culminating end-of-term *summative* measurement. Tests by themselves are not formative or summative (Popham 2011); it depends upon whether the test results are used for in-process analysis or outcome quality.

When looking at the instrument for gathering measures of learning itself, there are two achievement test-output categories to be considered: conceptual (commonly constructed and selected response) and performance (applied, task-oriented). Constructed response tests (short answer, essay) blur the boundaries between conceptual and performance. They are conceptual, and sometimes performance. Like stories, reports, or show-your-work problems, essays and extended-response test items are important forms of performance assessments (Airasian and Russell 2007). Educators' perspective on what constitutes performance varies, but the existence of three common characteristics prevails in identifying assessment as performance: (a) multiple evaluative criteria, (b) prespecified quality standards, and (c) judgmental appraisal (Popham 2011).

Careful design of learning measures in the context of Bloom's taxonomy of instructional objectives can further propagate performance when students are expected to recall and apply information in an actual task. In some cases, schools require student exhibitions, culminating projects, experiments, solving of realistic math problems, and various other demonstrations of competence (Slavin 2012).

3.2 *Informal and Formal Assessment*

When considering the possible data-gathering methods in the formative assessment realm, it is important to consider the purpose and weight of the check on learning. If the stakes are low, and the intent is to spur students to self-analyze their understanding of the material, informal snapshots of progress will suffice. In this

Table 7.1 Characteristics of informal and formal assessments. (Williams and Suen 1998)

Characteristic	Informal	Formal
Degree of freedom	Spontaneous assessment activities	Planned assessment activities
Flexibility	Flexible procedures/protocols	Prescribed procedures/protocols
Information	Depth of information	Precision of information
Objectivity	Subjective impressions	Objective measurement scores
Utility	Maximal informativeness	Maximal comparability
Bias	Subjective bias	Potential narrowed scope
Setting	Natural settings	Controlled settings
Inference	Broad inferences	Strong inferences

regard, a threaded or in-class discussion, a short written assignment, laboratories, or a weekly quiz can help illuminate students' connection with the course content. Through their work on these assignments, students discover weaknesses with the subject matter and have the chance to revisit content (Thorpe 1998).

The formative assessment event also delves into the formal assessment realm quite often. This typically materializes in the form of incremental, high-stakes exams (including midterms), presentations, and robust papers. The score, instructor feedback, and follow-up discussion of the formal check on learning act as a diagnostic tool for both the students and the instructor. For students, the corrective action may include increased study sessions and attendance of recitations. For the instructor, it may be analysis of the supplemental course materials, addition of recitations, and test-item analysis.

When combined in a concerted manner, informal and formal assessments provide meaningful information from which valid inferences can be made (Williams and Suen 1998). There is no unique formula for combining the two, but they must support each other in obtaining a reliable assessment picture of students and the instruction they complete. If the grading weight of informal checks on learning, such as quizzes, is kept relatively low, the in-process tools can act as preemptive indicators without overtly compromising students' final score. At the same time, students will maintain a shred of performance motivation while instructors have an indicator to redirect efforts toward points of student misunderstanding. Williams and Suen (1998) identified eight characteristic variances between informal and formal assessments as noted in Table 7.1.

3.3 Exploring Online Assessment Options

In online courses, the assessment options are impacted by the student population demographics. Although students take online courses frequently at the campus of which they are a resident due to schedule restraints, remote students pose the greatest assessment challenge. Nothing illustrates this point greater than the recent surge in MOOCs by highly ranked, large universities. Enrollments have hit 180,000 in a single course (Kolowich 2013b). Meeting the assessment needs of a population this large has only one option at this stage of the development, massive open online testing (MOOT).

On the other end of the enrollment spectrum, instructors have many more viable options for assessing students in capped regional online courses (CROCs). If a course is capped at 20–50, as is the case in many upper division courses, the assessment medium possibilities increase. Adding one or more teaching assistants provides added support for grading, but may pose reliability issues with regard to multireviewer subjectivity. The reduction of student numbers and the increase in instructor support allow opportunities for implementing more constructed written-response and performance-oriented activities as triangulated measures of learning performance. Additional details on MOOC and CROC assessment options will be discussed later.

3.4 Assessment Reliability

Reliability is commonly broken down into three variants (Popham 2011): (a) stability (test–retest), which refers to obtaining consistent results among different test occasions; (b) alternate form, which is the consistency of results among two or more different test forms (multiple-choice vs. essay for instance); and (c) internal consistency, referring to the way the test items in a single test function cohesively. Since this chapter is focusing on different test occasions (stability reliability) and various forms of assessment (alternate form reliability), internal consistency reliability will not be developed beyond supportive reference in the scope of this chapter. It is assumed that as a college instructor you have already covered the tenets of recommended test-design practices. The focus then will center on stability and alternate form reliability. For further tips on test design, refer to the works of Popham (2011), Airasian and Russell (2007), Tuckman (1999), Gronlund and Waugh (2009), and McMillan (2011).

When circumstances allow having a test proctored for online students, it will increase its reliability. Reliability in this instance refers to consistency over time. If an instructor were able to administer a test to the same group of students repeatedly over time, ideally the results would be the same. When a test is administered, aspects related to the test construction itself, the student, graders, and various circumstances surrounding its administration could cause the results to be inconsistent (Slavin 2012). One major factor that can affect the reliability of a nonproctored online exam is its equivalence to a take-home or open-book test. In a face-to-face or proctored scenario, the test taker is being monitored, albeit there are still many reports of unconventional test-taking practices during the monitored environment as well. With the take-home scenario used in standard F2F classes, it is possible to have supplemental in-person exams to provide triangulated measures and context. Furthermore, the instructor has a constant in-person engagement with the students, which can provide an opportunity for oral dialogue on the subject matter. In the fully online situation, however, it is difficult to gauge who or how many are working on the same exam. Given such a test, reliability is compromised.

4 Boosting Online Assessment Reliability

In the online testing environment, there are many techniques that one can implement to help reduce the reliability reduction: set the course management system to give random questions, place strict time limits on how long a student can spend completing the exam, make the exam available only within a short period of time (4 h on Wednesday for instance), allow only one or a few simultaneous users to complete a test at a time, alleviate moving backward in an exam, and request elaborate applied examples (not previously mentioned in the instruction) where applicable. Although this will make dishonest test taking more difficult, it will not foil determined collaborators or complex cheating schemes.

4.1 *Time and Resources*

The main issue with any of the alternatives for increasing online assessment reliability is time and resources. In a class of 20 students (i.e., a 20:1 student to instructor ratio), it is possible to implement assessment activities that require a person to evaluate each individual submission. When the student to instructor ratio expands, limited time inevitably forces an automated grading system that is confined to selected response instruments (choose an answer). This is not to say that selected response assessment is only a stand-in due to limited resource design. It is really the call of the instructor whether or not a selected response exam will adequately represent students' proficiency with the course subject matter. With selected-response instruments, it is still possible to write questions in a manner that leads students through a scenario cognitively speaking, have them perform a hands-on task (depending upon the testing environment), and select the appropriate response. Math story problems are a great example of this process. The term performance-oriented describes this process in a manner that approaches the performance-based process. Despite much criticism of selected-response exams, it is possible to tap students' higher-order thought processes through questions based upon varied taxonomies of educational objectives. There are times when knowledge-level understanding is necessary, and times when application or evaluation is necessary (Slavin 2012).

4.2 *In-Person and Virtual Proctoring*

The most obvious way to increase the likelihood that a student is the individual taking all the tests and the same person receiving the end-of-course grade is via identification verification and close in-person monitoring. Within this system, there are still instances of cheating that result in misrepresented outcomes or a decrease in stability reliability. The question to ponder is whether the take-home test or those completed online will have higher rates of academic dishonesty. In one

study (Watson and Sottile 2010), students reported equitable rates of cheating in F2F classes as compared to online classes, but 5.2% more had someone else give them answers during an online class quiz or test than in the F2F environment. Another question to ask is whether students who cheat would actually provide honest answers on a questionnaire intended to determine the rate of cheating.

Many instructors rest easier having exams take place under direct supervision. If you are an instructor who assesses students exclusively on essays or projects originating outside instructor or teaching assistant (TA) observance, questions of who and how many were involved may linger. Many colleges offer proctoring services where students can complete exams. A number of community, discipline-oriented, and private organizations provide this service as well including libraries and national testing services. Educational Testing Service (ETS) and Pearson Vue are two examples of private companies that administer many national certification exams.

Major issues pertaining to proctoring exist with regard to geographically place-bound students and the cost involved in setting up proctors. If monetary resources are not an issue, alternative options that mesh with today's technology footprint exist for offering proctoring services to remote students. Proctor U, Kryterion Inc., Pass My Exam, and Proctor Cam are a handful of the online proctoring service providers. They typically have an authentication process to determine the identity of the test taker. This includes ID verification, personal information verification, and real-time monitoring of the student via webcams and screen sharing.

5 Assessing Capped Regional Online Courses

In CROCs that are in fact capped with manageable numbers allowing individualized attention, instructors have many options for gathering assessment data. Multisource feedback for students provides triangulated measures of learning while allowing engagement that meets a variety of teaching and learning preferences.

5.1 Common Measures

Quizzes and exams have long been established as viable means of determining student-learning outcomes in college courses. One major variance between tests and other engagement activities is the reliance upon memory. It is generally believed that the information one remembers is what has been learned. In the online CMS environment, there are many options that will help increase the reliability of the quizzes and exams looking exclusively at the online delivery mechanisms. The following list highlights some of these items:

1. Limit the time to complete
2. Set the exam to be available for a short amount of time
3. Randomize question–response order
4. Randomize questions that each student receives
5. If multiple takes are allowed, set a minimum score for the first attempt before a second will be allowed
6. Conduct item analysis for each question (many CMSs calculate the data automatically)
7. Encourage student feedback after each question and exam
8. Preview each question closely if using test banks provided by textbook companies
9. Set CMS to allow only certain IP addresses
10. Categorize questions by response type
11. Provide clear directions for each section
12. Establish a specific protocol for glitches and resulting retakes

5.2 *Virtual Interaction*

Meeting with students online in videoconferencing is one way instructors assess students' content growth informally through interactive dialogue. Unfortunately, this poses a significant challenge due to the one-on-one time requirement, availability of videoconferencing technologies (hardware and software), and scheduling. Oftentimes, students indicate dissatisfaction when instructors of online courses offer them synchronously at scheduled times due to their time/place-bound circumstances. If a student is beyond a reasonable commuting distance or has set hours of employment, it is difficult to attend any scheduled class whether F2F or online. In the asynchronous delivery scenario, the time can be more forgiving, but the options for assessment are more limited.

5.3 *Virtual Presentations*

Group projects have the means to provide increased student understanding of content- and instructor-related advantages including multiple perspectives and pooled efforts (Young and Henquinet 2000). From an instructor's standpoint, presentations provide an alternative means for students to demonstrate their competency vested in a culminating course project (Arnold 2010). As a means to capture the presentation component of an F2F class in the online course delivery medium, major projects can be assigned with the presentation element at its core. The process of presenting acts as reinforcement for learning that will oftentimes motivate presenters toward adequate preparation and information grounding (Arnold 2010). Students are able to demonstrate meaningful, multidimensional tasks via this authentic assessment

(Montgomery 2002). This can be achieved through lecture-capture systems (Panopto, for instance) or through other computer-based presentation programs (Adobe Presenter, for instance).

5.4 Performance Pontification

Digital video editing is well suited for providing authentic, meaningful, reflective experiences for teachers (Calandra et al. 2009). If it is more pointed in its output with specific criteria, then it becomes a viable assessment tool. When constructed by the students who are being assessed, and as participants in the video, the instructor will be able to analyze the video for key levels of pontification pertaining to the course and assessment objectives as the following pontification assignment summary illustrates.

Arnold (2012) studied the feasibility of digital video editing through technology for educators courses, which were broken down into five modules, each with 3 weeks devoted to a specified theme. Given that the course is primarily for pre-service teachers, the focus was on pedagogy and using technology to support the standards-based subjects in the classroom. Theoretically and practically, teaching requires substantive merging of content, pedagogy, and technology knowledge (Roblyer and Doering 2012). Each of the modules had an overarching technology-based theme with multiple technologies addressed, substantial readings, academic content standards tie-in, pedagogical foundation, and emphasis on integration. During each module, students use and create comprehensive projects with multiple cloud and computer-based technologies while exploring an instructional delivery/e-learning concept such as podcasting. These key assignments throughout the semester have students expound upon their growth in the course content through various digital outputs that incorporate text, static images, audio, video, or a combination.

As a culminating activity near the end of the term, students were given a choice to either create a comprehensive digital story or write a paper on a subject of interest that would be covered in an elementary classroom and that supports a content area. If they choose the video, they may create it as either an individual or a group project (self-selected groups).

It may be anything from science (rocket propulsion for instance) to a social message (wash your hands frequently to reduce germs) or any other subject you would expect students of your favorite grade level to learn (look in the content standards for a grade level and subject of choice to identify a specific performance objective). In the spirit of the “reality” TV mash-ups (i.e., “Survivor,” for instance, where the program shows the tribes in action and then cuts away to an individual sharing his/her perspective on that action in an interview scenario), I would like you to interperse yourself into the video as the teacher giving your perspective regarding the use of technology in the learning/teaching process and with your chosen subject, while teaching the viewer about a chosen topic (i.e., the earth’s rotation/tilt and seasons). Students were encouraged to get kids involved if possible, and were given the latitude to complete it as a group project with other students.

In order to discern specific concepts critical to the learning outcomes, further detailed criteria were included. Some pertained to the technology skills, whereas others were targeting educational technology and integration with the elementary school subject-matter concepts. The following are abbreviated samples of measurable objectives included in the project:

1. Refer to and include specific educational technology supportive content from at least six journal article or textbook sources that were assigned during the term.
2. Devote about one-third of your video to talking about integrating technology into the classroom, and the remainder to teaching about a specific topic in a grade level and subject of choice. Be sure to combine them so it does not appear like two separate videos.
3. Include at least three motion video clips of yourself talking about integrating technology into teaching.
4. Include at least two separate audio clips of yourself talking about images, third-party motion video clips, technology integration, or explaining visual examples of the subject matter.
5. Connect with and identify multiple standards: information literacy, NETS*T, state academic content standards, and state educational technology standards for students.
6. Make the presence of each group member equitable and evident throughout the video.
7. Demonstrate competency with multiple technologies/processes: Movie Maker, Audacity, online file conversion, iTunes, YouTube, ID Tag Editor, synchronized and overlapping soundtrack, and narrations.
8. Effectively integrate still images, motion video, text slides, overlays, soundtracks, and narrations.
9. Include important elements of a presentation: introduction, body, and conclusion.

On a smaller scale, and in a similar manner, students were given a reading response assignment whereby they had to create an audio-narrated hypermedia presentation (PowerPoint) in which they identified key points made in the readings as text on the slides and discussed them in audio format. Having met with each student individually at the beginning of the semester in a videoconferencing site and requiring students to post audio introductions in their e-portfolios, the instructor was familiar with students' voices. Given such, it resulted in more personable assessment than written papers. When students devised audio-only pontifications, they were given reasoning and instruction on using a Wiki-embedded media player versus adding a more personable face to their audio compilations. This included embedding photos and personal information in their completed mp3 files using a program such as Mp3Tag, and embedding their audio files in Avatars (with the Voki program, for instance), which were in turn embedded in their Wiki e-portfolios.

A couple of drawbacks, especially pertaining to using multimedia in the online environment, include devising a systematically reliable audio/video evaluation means and the technological requirements for developing a video that represents one's content development. Validity is also important to this type of assessment.

Table 7.2 Performance pontification assessment criteria: TPACK

Pontification media assessment criteria		
Quality sources (peer-reviewed)	Duration	Technological detail
Encapsulates Ed Tech foundations	Optimum blend	Shared appropriately
Incorporates previous media covered	Self-presence	Submission deadline met
Reusable third-party media use/support	Standards-aligned	Includes K–12 students
Content source triangulation	Creativity	Teaches a topic/main point
Components of a <i>good</i> presentation	Media quality	Demonstrates design principles

Does the video allow an instructor to measure the conceptual knowledge that needs to be measured? Identifying the expected outcomes was not really a problem, but some students opted to read from scripts which can leave the evaluator wondering if the presenter is engaged or simply reading information that is not internalized.

5.5 *Evaluating Audio and Video*

From the grading perspective, there is still a disparity in reliability from one instructor to another. Multiple teachers grading the same essay paper will assign grades ranging from A to F with some teachers making few to no comments or marks on the papers, but instead just producing a grade (Brimi 2011). As noted above, there is a convergence of technology-use skills, technology integration with subject-matter propensity, and any given number of subtopics pertaining to educational technology covered in the course that must be weighed when evaluating a video produced by students of a technology for educators course. Time is a critical element in the analysis, but quality is the most decisive in determining if students are pontificating about the concepts covered in the course. In the scope of this analysis, most students were able to expound upon their chosen topic (a science lesson on volcanoes, for instance), but most commonly, underdeveloped their connection of the lesson topic to their use of technology to demonstrate it, or other examples of technology that would further support the teaching of the lesson. The next confounding factor that tended to affect students results, whether audio or multimedia, was the technology medium being used for the output (Table 7.2).

5.6 *Technological Factors*

Early in the course, the technology skillsets were more limiting to the quality of course concept-infused outputs than later in the term. Given such, the course was structured with less complex technological components in the beginning. Week-by-week new technologies are introduced. During the first 3-week module, students are introduced to relatively low-end technologies. During subsequent modules, as students’ efficacy climbs, they are directed toward more complex technological

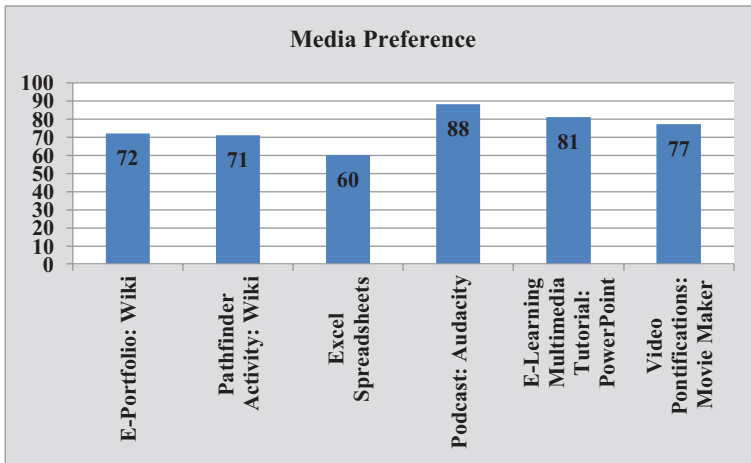


Fig. 7.2 Class rank-order preference of major projects

developments such as multitracked audio and video outputs using programs that balance user friendliness, effectiveness, and relatively free availability. These are characteristics that are likely to encourage preservice teachers to continue using technologies adequately when they transition to in-service status; a time that has many reeling from the steep learning and time commitment curve common during the first 2 years. Toward the end of the term, students in the technology for educators course are pushing some of the low-end technologies to their limit, which inevitably impacts their output and perception of technology.

5.7 Student Media Preferences

When students were asked if they preferred demonstrating content they have learned in the course through audio or video output (reading responses, interactive hypermedia, videos, etc.) over writing a paper on the same, 81% strongly agreed and 19% agreed. Given the number of technical glitches that were communicated during the term, it is curious that no students indicated preference for writing a pontification paper over creating the video. Perhaps, as Roblyer and Doering (2012) point out, technology can improve student motivation, attitude, and interest in learning.

In an end-of-course improvement evaluation, the instructor administers to students a rank-order question indicating that students prefer audio- and video-enriched technologies (see Fig. 7.2). The question only analyzed the larger project outputs without sub-analysis of the smaller technologies that most often fed into the larger projects. In the ranks for each project/media type, students identified Podcasting/Audacity (88) as their optimum output medium, with E-Learning/PPT (81) and Video Pontification/Movie Maker (77) close behind in that order.

Table 7.3 Student TPACK perception means

	Video		Podcast	
	Pre	Post	Pre	Post
Technology	8.67	7.81	8.86	7.62
Pedagogy	3.95	7.91	3.95	8.24
Content knowledge	4.86	8.48	5.14	8.10

In discussions with students, most seemed more enthusiastic about the outcome of their e-learning and video activities, although the higher rate of technology glitches, increased time commitment, and higher complexity level of the assignments associated with their development may have led to lower ranking than the podcast. Large file sizes, program freeze-ups, conversion to an iTunes U compatible mp4 file format, and student self-consciousness about presenting in the video were concerns voiced by students during the latter part of the term devoted to the multimedia projects.

5.8 Student Technology Perception

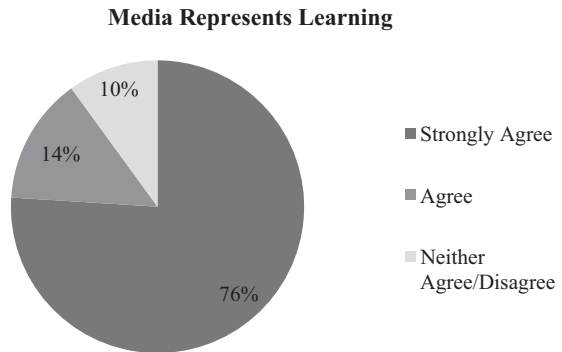
Students were given an additional questionnaire, the Technological Pedagogical Content Knowledge Alignment Perception Scale (TPACKAPS) at the beginning of the semester, and again at the end. Each student was asked to rate various components of the course (readings, discussions, papers, and media) on a scale from 1 to 10 for technology, pedagogy, and content knowledge emphasis (1 = none; 10 = primarily). Paired-samples *t*-tests were conducted to determine whether students’ perception of the course components varied upon completion of the course. The results indicated that the pre and post means (Table 7.3) varied significantly at the $p < 0.01$ level in students’ perceptions of videos and podcasts for technology, pedagogy, and content knowledge.

The positive correlation indicates that students perceive more technology-oriented focus in the beginning, but they perceive more pedagogy and content-knowledge focus after having substantial educational technology foundational development in conjunction with the media projects. When asked if they felt that media output represented their level of learning in the course with regard to technology, pedagogy, and content knowledge, 76% strongly agreed, 14% agreed, and 10% neither agreed nor disagreed (see Fig. 7.3).

5.9 Impact of Digital Output

The results of this study indicate that students prefer multimedia over other types of technology, and view digital video and audio as TPACK-rich media capable of demonstrating their competencies. Students perceive more pedagogical and content-knowledge potential in media postexperiential, and with proper attention called

Fig. 7.3 Students perception that multimedia represents and demonstrates their level of learning in a course



to the reasoning behind the processes being modeled in the technology for educators course. Since the students are preservice teachers, it is important to not only subject them to the processes but also to explain to them the scope of intentional teaching practices.

Given that the course utilized in this study is heavily infused with large doses of pedagogy and content-knowledge instruction in addition to the technology literacy skill development, a balanced TPACK approach is modeled for the students. Furthermore, students are challenged to create outputs that equitably merge each TPACK component. It is important to point out, however, that a number of challenges must be addressed during a media-intensive performance pontification project: purposeful media use; tech glitches; students must learn tech in addition to content; tech resource availability (software and hardware); file size; students must get to the point in the limited time (think about a TV show); multimedia principles must be covered; and students could read their script without fully engaging in the content.

Aside from an instructional and learning tool, video has been around for many years as a formative assessment, feedback, and planning tool. Common uses in this realm have included recording oneself giving a speech or presenting a student teaching lesson, real-time and postgame sports analysis, diagnosis of medical conditions or behaviors, pretest and posttest analyses of research subjects, law enforcement, and anything that requires a comparative stop action, archival capability. From an instructor's point of view, the digital audio and video output can offer a creative and visual dimension not represented in print. From the student perspective, whereby they are interjecting an audible or visual presence in the media, it is typically more common as a self-assessment tool.

Video tools are not uncommon as a means to teach content to others, even with the self in the visual mix. Through the use of audio and video, students are able to solidify their learning due to the increased cognitive processing needed to develop quality output. In addition, students will encounter added motivation due to the prospect of having a novel means to demonstrate their competency. Students are fascinated (and thus motivated) by such tasks as having an Avatar represent themselves with their own voice and remotely similar appearance. As a demonstration of

what one has learned based upon engagement in a course, digital audio and video in the online class environment is an underutilized and viable output. It warrants further analysis as an output tool not only in educational technology but also in less technology-focused disciplines as well.

Although students tend to like working with multimedia technologies in creating presentations as an alternative to prose output, the learning curve of the technologies adds more responsibilities to their shoulders. In a technology for educators course, it comes with the territory, but in other nontechnology related courses, low-tech options must be pursued. Lecture capture systems allow relatively low-tech alternatives, but the sacrifice is quality media presentation format and flexibility. It essentially becomes a talking head next to a PowerPoint presentation.

6 Assessing Massive Open Online Courses

Considering that the overall completion rate of MOOCs is noted to be in the 10% range (Kolowich 2013b) and that there is substantial interest in being able to offer these courses for credit, substantial advances in reliable assessment procedures need to be established. “Students who experience failure or disappointing grades carry negative emotions about their experience into their future learning” (Thorpe 1998, p. 268). The digital and other creative outputs discussed previously are simply not an option for courses enrolling tens to hundreds of thousands of students. Perhaps a categorization designation would be the first order of business to distinguish between credited open online courses (COOCs) and noncredited open online courses (NOOCs). The first could be tied to a highly weighted final examination proctored by colleges or by approved agencies that administer other certification exams. In this instance, it would act like the test-out option many disciplines maintain, except it would be supplemented by the remotely completed online summative assessments of questionable reliability. The second would be truly opening the doors of education to the world for anyone who purely wants to learn without credit or testing except as a means to reinforce learning via systematic quizzes.

7 Conclusion

Reliability issues in the online teaching front clearly need much more attention as we expand the online enrollments in courses. Offering proctored tests and test-out options is one solution, but not logistically feasible in many cases. Looking ahead, it would be beneficial to find out how many online courses are currently relying exclusively upon unmonitored online tests as the largest percentage of students’ final grades. There are elaborate organized schemes of online course assistance for students who are willing to pay, including test takers, paper writers, and entire course surrogate students. There are also plenty of honor-driven students, who complete their coursework via their own cognizance. They only want a better way to learn.

References

- Airasian, P. W., & Russell, M. K. (2007). *Classroom assessment: Concepts and application* (6th ed.). New York: McGraw-Hill.
- Arnold, S. D. (2010). *Student e-cooperative: Computer mediated group projects*. In the International Technology, Education, and Development Conference Proceedings CD (ISBN: 978-84-613-5538-9). Valencia, Spain: INTED.
- Arnold, S. D., & Moshchenko, M. (2009). Technology input versus input and output: Does it result in learning differences among elementary school students? In C. D. Maddux (Ed.), *Research highlights in technology and teacher education 2009* (pp. 1–9). Chesapeake: Society for Information Technology & Teacher Education.
- Arnold, S. D. (2012, October). Assessing online learning with digital audio and video. In Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012 (pp. 473–479). Chesapeake, VA: AACE.
- Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment Research & Evaluation*, 16(17), 1–12.
- Calandra, B., Brantley-Dias, L., Lee, J. K., & Fox, D. L. (2009). Using video editing to cultivate novice teachers' practice. *Journal of Research on Technology in Education*, 42(1), 73–94.
- Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
- Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of student achievement* (9th ed.). Upper Saddle River: Pearson Education Inc.
- Hunter, R. (2004). *Madeline Hunter's mastery teaching: Increasing instructional effectiveness in elementary and secondary schools*. Thousand Oaks: Corwin Press.
- Kolowich, S. (22 March 2013a). The minds behind the MOOCs. The chronicle of higher education. <http://chronicle.com>. Retrieved October 1, 2013.
- Kolowich, S. (8 April 2013b). Coursera takes a nuanced view of MOOC dropout rates. The chronicle of higher education. <http://chronicle.com>. Retrieved October 1, 2013
- McKenzie, G. R. (1979). Effects of questions and test-like events on achievement and on-task behavior in a classroom concept learning presentation. *Journal of Educational Research*, 72(6), 348–351.
- McMillan, J. H. (2011). *Classroom assessment: Principles and practices for effective standards-based instruction* (5th ed.). Boston: Pearson Education Inc.
- McTighe, J. (1996). What happens between assessments? *Educational Leadership*, 54(4), 6–12.
- Montgomery, K. (2002). Authentic tasks and rubrics: Going beyond traditional assessments in college teaching. *College Teaching*, 50(1), 34–39.
- National Survey of Student Engagement [NSSE]. (2010). *Major differences: Examining student engagement by field of study—annual results 2010*. Bloomington: Indiana University Center for Postsecondary Research.
- Popham, J. W. (2011). *Classroom assessment* (6th ed.). Boston: Pearson Education Inc.
- Roblyer, M. D., & Doering, A. H. (2012). *Integrating educational technology into teaching* (6th ed.). Boston: Pearson Education Inc.
- Roediger, H. L. III., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* Academic Press, 2011, Volume 55, pp. 1–36, ISSN 0079-7421, ISBN 9780123876911, <http://dx.doi.org/10.1016/B978-0-12-387691-1.00001-6>.
- Slavin, R. E. (2012). *Educational psychology: Theory and practice* (10th ed.). Boston: Pearson Education Inc.
- Struyven, K., Dochy, F., Janssens, S., Schelfhout, W., & Gielen, S. (2006). The overall effects of end-of-course assessment on student performance: A comparison between multiple choice testing, peer assessment, case-based assessment and portfolio assessment. *Studies in Educational Evaluation*, 32, 202–222.
- Swan, K., Schenker, J., Arnold, S., & Kuo, C.-L. (2007). Shaping online discussion: Assessment matters. *E-mentor*, 1 (18). http://www.e-mentor.edu.pl/_xml/wydanie/18/390.pdf. Retrieved October 1, 2014.

- Thorpe, M. (1998). Assessment and third generation distance education. *Distance Education, 19*(2), 265–286.
- Tuckman, B. W. (1999). *Conducting educational research* (5th Ed.). Belmont: Wadsworth Group.
- Watson, G., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? *Online Journal of Distance Administration, 13* (1). <http://www.westga.edu/~distance/ojdla/spring131/watson131.html>. Retrieved October 1, 2013.
- Williams, B. L., & Suen, H. K. (1998). Formal versus informal assessment methods. *American Journal of Health Behavior, 22*(4), 308–313.
- Young, C. B., & Henquinet, J. A. (2000). A conceptual framework for designing group projects. *Journal of Education for Business, 76*(1), 56–60.