Assessment of Active Learning

Kay Gibson and Carolyn M. Shaw

Introduction

Within the field of International Relations (IR), as well as across all academic disciplines, scholars have increasingly focused on the concept of assessment and its importance in measuring student learning and the effectiveness of teaching in the classroom. Beyond the classroom, assessment is used to evaluate curricula, departments, programs, and even universities as a whole. It is often the cornerstone of certification at different levels. This essay, however, has a narrower focus, looking specifically at assessment within the college classroom as it pertains to a variety of active-learning techniques.

The use of active-learning techniques has expanded in recent years and so have the different assessment methods used to evaluate these teaching techniques. The scholarship on assessment, however, is not as comprehensively developed as the teaching methods themselves. In order to provide a richer context, we will draw not only on literature within International Relations, but also from other disciplines including education. We recognize that assessments can be done for traditional, lecture-based classes as well as classes that incorporate active learning, but the focus of this essay is on assessing active learning, which includes case studies, simulations, games and role-play, small group discussions, collaborative group projects, and service-learning as some of the most widely used exercises.

The purpose of this essay is not necessarily to advocate for the use of active-learning techniques but to explore effective student assessment methods for various types of active learning. It is important to capture the state of assessment in the field so that further progress can be made in advancing assessment techniques and effective teaching. This essay begins with an overview of the historical development of assessment tools in general, followed by the development of active-learning assessment tools, as found in educational and discipline-specific literature. We then identify some of the best practices in the assessment of active learning by asking what, why, when and how instructors choose to assess. This section includes discussion of learning objectives, formative and summative assessments, and some of the essential components of assessment. Next we provide specific examples of assessment techniques and measurement tools that can be used in different active-learning contexts. We conclude by identifying several areas for future research in the area of assessment, noting that there is still much to be done in terms of scholarly review.
**Historical Assessment Tools**

Learning assessment tools have evolved over time, just as learning objectives and teaching techniques have. For many years, learning objectives were focused on content knowledge and factual information. Lecturing was the teaching strategy most often employed in collegiate classes, with written exams used for assessment. Test items took on various forms such as true/false, multiple choice, short essay and sentence completion; but the assessment was structured as an exam. The type of content being conveyed contributed to the high use of exams as an assessment tool with a large portion of content related to facts and understanding of discipline knowledge. Assessments required “regurgitation” of the facts and, for the most part, only a basic application of knowledge. The one-way transmission of knowledge through lectures, and assessment by exams, required little critical thought or analysis, synthesis or evaluation of the content (Bloom et al., 1956). What was required was memorization of the content to pass the exams (Wiggins and McTighe 1998; Prince 2004).

In the 1960s, educators began to recognize that new knowledge was developing at an increasing rate and the charge to instructors to be the conveyors of knowledge was impractical and ultimately unrealistic. Professional educators began to examine other types of content that needed to be taught in order to enable students to effectively deal with the expanding knowledge base. In the field of education, the teaching of content that involved skills, processes and attitudes began to take on new importance. Students needed a set of processes and skills to handle the barrage of information (Major and Palmer 2001). Active-learning experiences that would allow students opportunities to learn through participation and practice using skills and processes took on a more prominent role in classroom instruction. Teaching skills by employing them rather than by simply listening to a lecture about them was a more effective teaching method. For example, lecturing about how to problem solve in a collaborative setting is not as efficient as engaging in collaborative problem solving as active learning in the classroom.

With the shift in learning objectives that were more focused on the development of skills and processes, new assessment techniques were required to be developed to determine if these skills were being taught effectively through new active-learning techniques. However, these newer assessment methods, as well as scholarly reviews of them, have been slow to develop.

**Development of Active-Learning Assessment Tools**

Early literature on active learning devoted much effort to trying to prove that active-learning techniques were more (or less) effective than traditional lecture-based teaching (Bloomfield and Padelford 1959; Guetzkow et al. 1963). There are still some of these articles today, but the debate has largely become more nuanced in recent years, with scholars recognizing that no single teaching technique is superior in teaching all types of content and skills. There is a growing consensus that the technique to be adopted should be selected based on the specific set of learning objectives. Active-learning methods seem particularly well suited to learning outcomes related to affective learning and the teaching of practical (or professional) skills, such as negotiation, mediation, bargaining, and consensus building. Dorn (1989) argues that active-learning techniques are more effective at teaching “critical thinking and problem-solving skills.” The support for this type of argument is discussed in more detail below.

Despite the recognition that different objectives can best be taught using a variety of methods, faculty still often use written exams to determine what students have learned through active-learning exercises. An example of one assessment method that works for knowledge acquisition, but not for skill acquisition is the pretest/posttest. Students are asked a set of questions before and after the instruction unit. Pretest/posttests, can be used to determine whether or not students have the basic knowledge that will equip them to learn and understand more complex knowledge and skills related to a discipline or field. It is difficult, however, to identify what skills and advanced processes students have learned through a standardized, written exam. For example, measuring how adept a student is at negotiation is difficult to quantify, and tends to be subjective. Sometimes oral examinations are used, providing students with a different medium to express what they have learned. This can be especially useful for students who have different learning styles and strengths. Mastery of various skills and processes, however, is often best assessed through demonstration rather than through written or oral explanation. Written briefs, oral arguments, and debriefing discussions provide additional points of measurement beyond standard tests, with evidence for instructors that students are learning the material.

The literature below explores several different claims about active learning, and then examines how well they are supported within the literature on assessment. Unfortunately, the assessment literature has not been robust in scientifically proving the added value of active-learning techniques.

There are many articles that seek to illustrate the positive results that stem from active-learning techniques. Krain and Lantis
(Jensen 1998), enhance student problem-solving skills (Bransford et al. 1989; Lieux 1996), and increase the retention rates (Stice 1987; Schacter 1996; Silberman 1996; Hertel and Millis 2002). They also aid students in understanding abstract concepts (Pace et al. 1990; Smith and Boyer 1996). Despite these claims, however, Raymond (2010) emphasizes a surprising lack of systematic empirical evidence. Several recent studies note that most claims are based on subjective impressions of the instructor or students (Shellman and Turan 2006; Powner and Allendoerfer 2008). There are many examples of anecdotal support for the effectiveness of active-learning techniques. These can take a variety of forms (Newmann and Twigg 2000). Many are done through debriefing sessions. Chasek (2005) invites students to sit in a circle and discuss the challenges of negotiation and organizational structural reform following a Security Council simulation. Students “admitted that the simulation gave them an entirely new perspective on the problems encountered in changing the membership and modalities of the Security Council” (Chasek 2005). Switky (2004) recommends letting students work in small groups to summarize the exercise’s important lessons and then share these with the whole class. Lessons that emerge are often comparisons of how the simulation and real life differ. Zeff (2003) notes that outspoken students can dominate oral debriefings and recommends a written component. In the student surveys of her European Council simulation, students state that “they are more actively engaged in learning the class material because they have to defend their countries’ positions to the other participants” and that the simulation “helped them to appreciate the complexities of decision making in the EU [European Union]” (Zeff 2003).

Another option to determine the value of an active-learning exercise is through the use of standard course evaluations completed by students at the end of the semester. A study by Oros (2007) compares the data from six courses that employed structured class discussions and nine courses that did not. Evaluation scores were higher in the upper division classes, where debates were used, than in the classes without debates. The data, however, were drawn from a single survey question: “Readings and assignments were a valuable part of this class” (Oros 2007). The hand-written evaluation comments reflected these numbers, with students noting that structured classroom debates “helped me form opinions on various important topics,” and “allowed me to write a more knowledgeable paper and understand information in a more informed manner” (2007). Some surveys include both students and tutors, looking at their perceptions of the learning experience. A study by de Freitas (2006) notes that student comments validated the instructor perception that students have an increased motivation when learning through the use of games. The outcomes of these various assessment and debriefing methods are generally positive. While these assessments are useful, they are not particularly rigorous, however.

Studies that employ more rigorous methods produce mixed results. Krain and Lantis (2006) designed a careful study with two classes and two simulations. One class served as the experimental group, the other as a control group, and then these roles were reversed. Students were given pre- and posttests on acquisition of knowledge and were also asked questions about their perceptions of their learning. Both lecture and simulation methods appeared to give students confidence in their knowledge acquisition (this was also reflected in the objective test scores), but only those who engaged in the simulation felt that they had also gained a better understanding of actors’ preferences and the complexity of the processes involved (Krain and Lantis 2006). The authors conclude that “overall, this experience suggests that while each technique produces learning, neither produces greater knowledge gains than the other […] [although] the active–learning technique and the lecture/discussion approach may affect student learning in different ways” (Krain and Lantis 2006). Powner and Allendoerfer (2008) also conducted a study with a control group. They did see an increase in test scores of those students who engaged in a role–play exercise, but this did not extend to an overall improved class performance. Raymond (2010), on the other hand, conducted a study of students in seven different sections of a course, Introduction to International Relations, and found there was no difference in the academic performance between the students who participated in the role–play exercise and those who did not.

The lack of conclusive literature on assessment of students in active learning is likely linked to the challenges involved in scientifically measuring the benefits of such exercises. Raymond (2010) notes a number of research design issues that must be considered. One of the most important is having a control group and a test group to show a demonstrated effect of the exercise that is being assessed. Another key factor is having a large enough sample (large n) to make the evaluation data meaningful. When evaluating the claim of greater retention of materials, a longer–term study is often needed that extends beyond the end of the term. An additional challenge includes posing questions that adequately capture the knowledge acquired by the students. If the questions are not well crafted, they may not reveal the learning that has occurred (i.e., students may have learned something more than what they were asked about). There are also unique student characteristics to take into account that may affect performance. Unless a study includes data on class status (Fr, Soph, Jr, Sr), degree major, and general academic ability (GPA) – to name a few – the results of assessing a particular active–learning exercise may not tell the full story (Wheeler 2006). These student characteristics can also have a varying impact depending on the topic of the exercises. Groves, Warren, and Witschger (1996) describe a multi–course study on social inequalities. White students expressed different lessons that were learned than minority students. Even environmental factors might account for variance in performance, such as whether a class is taught in the morning or afternoon, or in the spring or fall semester.
These are all things to consider in a scientific study, but are often beyond the control of instructors. While it may be possible to have a control group for a large introductory class with multiple sections, it is more difficult for specialized upper division classes, or in situations where different colleagues teach the same sections (another variable to try to control). The same challenge exists for having a large enough sample size. While it is possible to engage in longer-term studies, it is often difficult to track students beyond the semester in which they are enrolled in a particular course. These challenges are of particular concern if one is trying to compare the value of an active-learning exercise to a more traditional teaching method. If an instructor is more interested in documenting the learning of individual students throughout a course, there are many ways to achieve this without worrying about sample size and control groups.

**Best Practices**

A scholarly understanding of assessment within IR has gradually evolved over the years, with significant insights coming from other disciplines, particularly education. Taking into account some of the challenges of assessment noted above, this section examines various assessment techniques that can be used with active learning including written and oral debriefing, observations, peer- and self-assessment, and presentations. This section also explores some best practices noted in the literature. In order for assessment to be done well, there are a number of questions that need to be addressed, all of which are interconnected. First, instructors must consider what they are assessing. This relates to learning objectives, both large and small, for a course or a specific exercise. Second, instructors should clarify why they are assessing and what benefits will derive from the process. Next, instructors need to consider when they will conduct assessments, during or after the learning process. Finally, instructors need to specifically address how they will design solid assessments of active learning to best suit their needs.

**What Are We Assessing?**

Assessment is a more complex process than many people realize, especially when what is being assessed involves content that is more than simply factual knowledge. In order to effectively assess student learning, instructors must first be clear about what the purpose (learning objective) for the learning exercise or activity is. Learning objectives drive both the selection of active-learning strategies and the type of assessment tools used to measure student learning and to evaluate the instructor’s curriculum design and implementation of instruction (Runté 2004; Victorian Curriculum and Assessment Authority 2007a). Over 50 years ago, curriculum theorist Ralph Tyler (1949) identified four elements of curriculum: objectives (or learning outcomes), content, strategies and assessment. Tyler saw the development of curriculum as sequential and linear; in other words, the instructor designed a curriculum plan by determining the learning objectives, selecting the content related to the objectives, choosing ways in which to teach the objectives (strategies), and assessing the results to improve future reiterations of the curriculum. Today, these elements are developed in a more interactive manner since all elements are interconnected and as such need to be considered together as assessment is being designed (Wiggins and McTighe 1998). In particular, the objectives will determine, to a great extent, the type of assessment that should be used to effectively assess the learning process. In a similar way, the content will prescribe the strategies needed to effectively teach the content. Therefore, it bears repeating that, in order to assess active learning, instructors need to first consider the objectives proposed for the learning activities. Runté notes that active learning demands active assessment (2004).

Some instructors adopt basic active-learning strategies such as small discussion groups with the simple goal of increasing student engagement. There is some support within the literature regarding the effectiveness of such exercises for this purpose (Prince 2004). Instructors, however, often have more multifaceted learning objectives for students beyond increased levels of engagement. Greenblat (1973) identifies several different categories of learning objectives that can be aided by using active-learning methods: cognitive learning (gain factual information, concrete examples of abstract concepts, analytical skills, procedural experience, and decision–making skills), affective learning (changed perspectives and orientations toward various public and world issues, increased empathy for others, and greater insights into challenges faced by others), and increased self–awareness and a greater sense of personal efficacy. The variety of objectives warrants a variety of teaching strategies and different types of assessment. Active learning not only provides a number of strategy possibilities but also has embedded within it a variety of assessment opportunities and possibilities.

If the learning objective consists of content knowledge, then written exams are a logical and effective method of assessment. A learning objective concerned with knowledge and facts is easier to assess because there is usually one correct answer. But assessment becomes more complex and challenging when the learning objective involves acquisition of new skills and processes that incorporate higher order thinking skills such as analysis, synthesis, and evaluation (Bloom et al. 1956). Skills and processes can be performed with different degrees of mastery. Instructors must analyze the skill or process (often called task analysis) to determine what the essential components of the activity are and what behavior the student must demonstrate to show mastery. Instructors need to be clear on whether the primary goal is student acquisition of content
knowledge or something beyond that. They should also recognize that any particular lesson or exercise cannot achieve all of
the learning objectives noted above. Setting clear, attainable objectives is important. Once the objectives are identified, an
instructor should also consider the purpose behind conducting assessments.

Why Are We Assessing?

There are many motivations behind choosing to do assessments in the classroom. The most obvious one is to evaluate
student performance in order to give a grade. This is true for lecture as well as non–lecture based courses. A second
motivation is to provide useful feedback to students so they can improve their performance (Steadman 1998; Chappuis and
have the capacity to improve learning. Students realize that they need to determine what is important and therefore what will
be assessed so that they can direct their efforts toward that content (Gibbs and Stimpson 2004–5).

Similarly, assessments can provide feedback to instructors about student learning so that they can modify their teaching
methods to clarify misunderstandings and increase the efficiency of the learning environment (Steadman 1998; Boston 2002;
Chappuis and Stiggins 2002; Guskey 2003). The use of assessment early in the study of a topic can help the instructor to
determine ability levels, learning styles, and areas of interest. With this information, the instructor can create activities that
engage students at appropriate levels of challenge, in ways they can learn more easily about things they have a real interest in
(Tomlinson 2007–8). Sound assessment is not separate from instruction. Because students learn at different rates and
through different teaching styles, assessment reveals these distinctions to the instructor (Tomlinson 2007–8; National
Capital Language Resource Center 2004). Using assessment to determine the effectiveness of the instruction is critical to
maintaining a high quality of teaching (Chickering and Gamson 1987).

Finally, assessment can be used to demonstrate the effectiveness of a particular teaching technique, in which case the active-
learning technique might be compared to a traditional instruction method, or contrasted with an alternate active–learning
strategy (Dunlap and Grabinger 1996; Prince 2004).

Assessment can be a valuable practice for all of these reasons. Beyond simply providing a grade, well–designed assessment
tasks will create opportunities for students to demonstrate the knowledge, skills, understandings, and attitudes that they
have learned. Assessment criteria and tasks can also act as a motivation tool, by allowing the students and others to
recognize their accomplishments (Victorian Curriculum and Assessment Authority, 2007b).

When Are We Assessing?

There are several different times in the learning process that assessment can take place. Assessments that are conducted
during the learning process itself, allowing for modifications in both teaching and learning, are formative assessments. Those
that occur after a learning process is complete are summative assessments. The choice of timing for assessment depends on
what an instructor plans to do with the information. Summative assessments are often used simply to provide points toward a
final grade and to determine the final level of student achievement. Formative assessments are used when instructors want to
modify their teaching methods based on student learning, or when they want to allow students to modify their learning based
on performance feedback.

Formative assessment is diagnostic in nature as it provides information about student learning that can be used to adjust and
improve teaching and learning to better meet the needs of students (Black and Wiliam 1998). Chickering and Gamson (1987)
maintain that students need frequent opportunities to receive feedback about how well they are learning and about ways that
might improve their learning. Thus, active learning that has formative assessment throughout a unit of study can provide
these opportunities for feedback and enhancement of learning. Assessment is not an end, but a beginning to better
instruction (Tomlinson 2007–8).

Formative assessment can be achieved through observation of classroom activities such as discussions, student
presentations, self– and peer–assessments, and small group work, but can also be done through pop quizzes, homework,
and formal tests that are positioned within the instruction of a unit rather than at the end (Boston 2002). Obviously, formative
assessment does not often occur in traditional, lecture–based teaching where communication is didactic and one–way, with
little or no input from the learners. Active–learning techniques, however, provide more opportunities for formative
assessment to take place as the instructor observes student performance and modifies the learning experience accordingly.
For example, student preparation for a role–play exercise might involve writing a preparatory paper prior to the simulation.
The instructor is able to determine from these papers whether the students have the necessary background to successfully
engage in the exercise or not. If further information is needed, it can be provided in order to maximize the learning from the
role–play exercise. Similarly, if an exercise extends over several days, it can be modified following each session to make sure
students are achieving the identified learning objectives.
Summative assessment is conducted at certain designated times in order to determine if students have attained the intended learning objectives at the end of the unit or course (Garrison and Ehringhaus n.d.; Popham 2008). This type of assessment is most often associated with unit tests, mid-term exams, semester exams, and final presentations or projects. Unlike data from formative assessments, generally speaking, summative assessment data are not used to adjust instruction during learning. When students receive summative assessment data, it is too late to modify their learning behaviors if they are not making satisfactory progress. Summative assessments are more likely to be used to determine student scores and grades.

Both formative and summative assessments can play a valuable role in the assessment of active–learning objectives. Summative assessments provide an efficient way to gain data to determine whether or not students have met course goals and achieved the learning outcomes at the conclusion of units of study and at the end of a semester. Formative assessments used during the learning process are more immediate and allow for modification of teaching and learning during the instructional process. Not only do both types of assessment data reveal information about student learning, the data also allow an instructor to productively reflect on the effectiveness of the curriculum and program including the choice of instructional active–learning strategies (Brown 2004–5; Fisher and Frey 2007). The best practice of assessment for students participating in active–learning tasks will employ both types of assessment – formative and summative.

**How Are We Assessing?**

Having explored the literature regarding the content, purpose, and timing of assessment, we turn now to the practical aspects of how to conduct well–designed assessments. We begin by noting the essential components of sound assessment found in the literature, whether used with traditional instruction strategies or in the context of active learning. A discussion of specific assessment techniques and measurement tools follows in the next section.

One of the most important components of sound assessment is for instructors not only to clearly identify their learning objectives (as noted above), but also to make them explicit to their students (Runté 2004; Tomlinson 2007–8). Students are more engaged and motivated when they have a clear understanding of what is expected of them and how they will be assessed. This is particularly true for non–traditional exercises such as simulations or other group activities where students need a clear set of expectations for performance and assessment. One way in which these expectations can be established is by actively involving both students and instructors. Where possible, assessments should be developed with student input so that they feel ownership and see purpose in the topic and assessment (Runté 2004). Thorpe (2000) proposes the use of self–assessment that requires students to articulate their own criteria and then assess their performance against the criteria. Even if students are not directly involved in designing assessment measures, instructors should provide a clear assessment rubric.

Another important component of sound assessment is the use of multiple and varied measures for the most reliable and valid assessment data (Angelo and Cross 1993; Thorpe 2000; Brown 2004–5; Tomlinson 2007–8). By its nature, active learning provides numerous opportunities to use a variety of assessment methods. Some of these techniques are often combined. For example, a peer–assessment and an oral debriefing session may be used after a student or group of students has made a presentation. In this instance, multiple assessments will help to create a more holistic and valid assessment of the presentation. These multiple measures, however, also need to be efficient in terms of learning and teaching time (Brown 2004–5). One way to make assessment efficient is to integrate it into learning. Formative assessment activities conducted during a unit of study using active–learning strategies can support, as well as extend, student learning (Runté 2004; Tomlinson 2007–8). By integrating formative assessment into the learning activity/task, opportunity is created for continued student learning even as the instructor is assessing students.

On a similar note, assessments should be designed, to the extent possible, to result in authentic, real–life products to which the learner can relate. In other words, the assessment results in something that would be required in the work place or “real world.” For example, students engaged in a mock Security Council simulation would write resolutions just like the real Security Council does in order to address crises around the world. This “real life” connection increases motivation and deep learning (Tomlinson 2007–8). Such real–life assessments allow the instructor to assess knowledge, skills, and behaviors in an integrated fashion instead of as a discrete criterion that helps to minimize unnecessary duplication of assessment tasks. It also creates assessment situations that more clearly reflect how students actually learn and how they transfer that learning to new and different contexts (Victorian Curriculum and Assessment Authority, 2007b).

A final component is that sound assessment should be an ongoing process that occurs at various points throughout the instruction (Chickering and Gamson 1987; Angelo and Cross 1993). Instructors should rely less on comprehensive exams and provide more periodic assessments with useful and timely feedback (Hattie 1987; Black and Wiliam 1998; Greer 2001). Formative assessment provides information that shapes teaching, in terms of sequencing of study, planning instruction and learning time, when to re–teach or move ahead, when to explain or demonstrate in another way.
Assessment Techniques and Measurement Tools

The above discussion of what constitutes sound assessment suggests that there are various types of assessment that are well suited to instruction employing active-learning strategies. In this section, we focus on assessment techniques that have been found to be effective when an instructor implements active learning. These assessment techniques include: written and oral debriefing, observations, peer- and self-assessment, and presentations and demonstrations. In addition, we look at several different measurement tools for recording the assessment data, including checklists and surveys. We conclude with a discussion of rubrics.

As was previously stated in this essay, traditional techniques of assessment are often inappropriate because of the very nature of active–learning experiences. Exams cannot assess reliably the acquisition of skills and processes that are often the desired learning objectives of active learning and teaching (Major 1999; Major and Palmer 2001). To ensure that any assessment is appropriate, focus should be on the learning objectives, with clearly articulated student actions and behaviors that demonstrate attainment of the objectives (Tomlinson 2007-8). Assessment techniques for active learning tend to be linked to student performance and products that are observed or judged, often in class, by the instructor and analyzed according to a set of criteria to determine if the student has achieved the learning objectives. Creating clear criteria is important and can be achieved with the construction of an assessment rubric that is specific to the learning objectives of the curriculum being assessed.

Debriefing

Debriefing can be used for assessment on an individual basis or collectively. If conducted collectively as a group activity, then the oral debriefing becomes more of a learning tool than an individual assessment – students learn further from hearing the comments of their peers, contributing their own responses, and comparing/contrasting the differences. Debriefings can be instructor- or peer-led. Instructor-led debriefing leaves the instructor in control, allowing him or her to emphasize the desired points. Student-led debriefings may take a bit longer, but may allow for a wider variety of responses and unanticipated observations.

Oral debriefing is especially useful in role–play activities and simulations. Depending on the structure of the role–play activity, students can complete the activity with an oral debriefing or with an individually written reflection followed by an oral debriefing. Debriefing is useful because at the completion of the simulation, students may tend to focus on “who won,” and a debriefing assessment activity can bring the discussion back to examination of the process or procedure. This emphasizes the higher learning skills that are often the learning objectives in active–learning exercises rather than simply content knowledge. The debriefing is a critical part of a role–play because it provides the opportunity for the instructor to ask students to discuss, to reason, to draw conclusions, and to link abstract concepts to practical experiences. It also allows the class to more fully recognize the dynamics that were at work during the exercise. While students are carrying out their own tasks, they may be oblivious to the incentives and actions of other students. It is often hard for the instructor to follow all of the interactions that are occurring within an exercise because of multiple conversations that occur at the same time, or because some actors will act in secret (outside the classroom, or through written communication). It is good to recognize, however, that oral debriefing cannot give the instructor feedback about each individual student’s learning since every student will not respond to every discussion question posed by the instructor. Thus it is useful to combine debriefings with an additional mode of assessment.

Observations

Observations are a powerful mode of gathering ongoing data of students’ learning (Victorian Curriculum and Assessment Authority, 2007a). They can take place in a variety of settings, across many activities, and employ a number of different tools to record information including checklists, anecdotal records, frequency count tables, and rubrics. When planning to observe students, instructors should consider whom they want to observe, what to observe, and how to evaluate and document what they see. Instructors may choose to select smaller groups of students over longer periods of time and focus on particular skills or knowledge to be observed.

An observational checklist is a common measurement tool that can be used to record observed student behaviors indicating progress toward achievement of the learning objectives. Checklists can be used to assess a student’s progress at a designated point in time or to assess the progress over time (National Capital Language Resource Center, 2004). These are based more on interactions than on outcomes. Checklists can be useful for assessing classroom tasks and activities because they are easy to construct and use, and can be easily written to align with the task(s) to be observed. When constructing such a checklist (see Figure 1), it is wise, however, to keep the number of criteria to no more than 10. Otherwise, the checklist becomes cumbersome physically as well as visually.
The checklist may also include role-specific criteria for a particular exercise. For example, if a student were playing the role of a mediator in a conflict situation, the observation criteria might include whether the student used positive or negative inducements to get the parties to compromise, whether the student maintained a neutral stance or took sides with one of the parties, and whether the student was able to identify commonalities to unite the parties. The end goal is to get the conflicting parties to engage in constructive dialogue and reach a negotiated settlement, but whether or not they reach a settlement is only part of the exercise, the negotiation process itself and how it is conducted are the key parts of the exercise.

By examining the observation criteria, one can ascertain with a fair degree of accuracy what the learning objectives are for the collaborative activity. When the criterion "respects the opinions of others" is considered, it becomes obvious that one of the objectives is that "the student will demonstrate respect for others and their opinions in verbal interactions with other group/team members." This mirroring of wording exists when learning objectives are used to write the assessment criteria and ensures that the instructor's observations are assessing what is to be learned. Sometimes the observation checklist can be used explicitly in a debriefing. The observation criteria may not be based on "right" or "wrong" behaviors stemming from the learning objectives, but on different approaches taken in the exercise. When a note is made of these different approaches, this can be used to reflect on the final outcomes during the debriefing session. For example, if the parties did not reach a final agreement and the mediator did not use positive inducements with the conflicting parties, then the students might conclude that following a different approach could yield different results. The contrast and comparisons become even more interesting if one group of students took a different approach from another group and had either similar or different results.

The observational checklist helps capture what is occurring during the exercise itself and allows for assessment and further reflection afterwards.

Peer-Assessment

Peer-assessment refers to the assessment of students by other students and is another mode of assessment associated with the use of active learning. Instructors ask students to assess each other, particularly when working in a group (Baker, 2008). As with observations, peer-assessments are based on the behaviors of others and use some type of recording tool like a checklist, inventory, rating scale, or rubric developed from the learning objectives to guide the assessment. The most common types of peer-assessments are rating scales and a single score method (Baker, 2008).

Despite occasional student discomfort with peer-assessment, it has been found to produce peer judgments that are comparable to those of the instructor (Patri, 2002). Matsuno (2009) found that peer-assessors were less biased than self- and instructor-assessors and were internally consistent in their ratings. Peer-assessment can benefit the learning for the student receiving the feedback as well as the student conducting the assessment (Patri, 2002). It promotes reflection and higher order thinking as students develop skills in evaluating and justifying the decisions they make (Victorian Curriculum and Assessment Authority, 2007a). In addition, peer-assessment can be a “real world” exercise, as supervisors are required to evaluate their employees in professional work settings. Peer- and self-assessments are often undertaken together because assessing the work of other students helps students to reflect on their own work and learn more effectively (Cheng and Warren, 2005; Matsuno, 2009).
The quality of peer-assessment is dependent on the quality of the tools, support given by the instructor, guiding questions that are asked, and the consistency of engagement in the assessment process (Victorian Curriculum and Assessment Authority, 2007a). An example of a well-developed assessment tool comes from Cheng and Warren (2005) (see Figure 2). Though used in an English language program for undergraduate engineering students, the tool could be used for oral presentations in any university course with little or no revision.

![Figure 2](Oral presentation assessment tool used in an English language learning class by undergraduate engineering students (adapted from Cheng and Warren, 2005))

**Self-Assessment**

Self-assessment has the potential to be a powerful mode of assessment that encourages students' self-regulation of their learning and setting of goals for self-improvement. Before engaging in reflection and self-assessment, students need to be familiar with assessment criteria that are clearly stated and with the objectives of the unit being taught (Topping 1998; Thorpe 2000; Lindblom-ylänne et al. 2006). Using the criteria, students need to examine their work, consider what they have achieved and think about what still needs further development. Like peer-assessment, the quality of the self-assessment depends on the clarity of the criteria, the detail of descriptions in assessment tools, instructor support during the process through guiding questions and opportunity for student questions, instructor modeling, and regular opportunities to practice self-assessment through reflection (Thorpe 2000). Peer- and self-assessments are most effective when they are embedded into the learning in the unit and students are provided with the opportunity to learn from their mistakes in a non-threatening environment.

Often, self-assessment is achieved through the use of a reflective journal, written with the assistance of a set of guiding questions. Journaling offers another way to encourage understanding of oneself as well as understanding of concepts and their application to life experiences outside the classroom (Langer 2002; Park 2003; Chabon and Lee-Wilkerson 2006). Self-assessment through journaling works well in the context of student internships and other experiential learning opportunities. Self-assessment can also be accomplished through data collection tools similar to those used for peer-assessment such as inventories, rating scales, and rubrics. Effective self-assessment helps focus students on their individual strengths and attitudes, analyze their progress, and set goals for subsequent learning. Self-assessment through reflective practice is an important life-skill to develop because many work contexts utilize self-appraisal in their employee review process (Baker, 2008). Self-assessment is a meta-cognitive activity and an essential component of student reflection that causes the student to become more conscious of how they learn, the quality of their work, and the purpose of the learning experience (Carson and Fisher 2006; Lindblom-ylänne et al. 2006).
Presentations and Demonstrations

Presentations and demonstrations are both authentic assessment techniques that are valuable in active-learning environments. They provide students with the opportunity to develop “key, transferable skills […] and to make the connection between their learning and real world learning contexts” (Victorian Curriculum and Assessment Authority, 2007a). By verbalizing in presentations, and making explicit their learning through demonstrations, students are applying a variety of skills related to meaningful, everyday situations. For example, students may demonstrate leadership or management skills by running a mock meeting. Measurement tools that are particularly effective for the assessment of presentations and demonstrations are similar to those discussed for debriefing, peer-assessment, self-assessment, and reflection. It is critical, however, that the assessment tool is designed in an easy to use format, because presentations and demonstrations only last for a finite amount of time and cannot be reexamined to complete the assessment tool.

Student Surveys

Student surveys are an effective way to assess all aspects of active learning. In particular, surveys are useful in assessing affective learning objectives such as those related to changes in students’ attitudes, perceptions or perspectives (Schulte and Carter, 2004; Wee et al. 2004; Lingefjard and Holmquist 2005). Surveys can also provide direct feedback from students on the effectiveness of the instructor’s teaching and the students’ satisfaction with the course (Czaja and Blair 2004; Walker and Kelly 2007; Combs et al. 2008; Koc and Bakir 2010). A survey is an assessment mode as well as a measurement tool. Because surveys are often used to assess attitudes, a level of agreement or disagreement is generally measured. For example, a typical five–level Likert item might be: 1–Strongly disagree, 2–Disagree, 3–Neither agree nor disagree, 4–Agree, and 5–Strongly agree. Open-ended (non-structured) questions and closed-ended (structured or fixed response) questions are also used in surveys. An open–ended question might be: “What did you like best about the Peace Summit Simulation?” If the survey uses closed–ended questions, then choices would be provided and the students would be required to respond within the confines of those choices. For instance:

What did you like best about the Peace Summit Simulation?

   a. Being able to come up with a solution to a challenging global issue.
   b. Getting to learn more about the foreign policy preferences of country X.
   c. Being able to make speeches and draft diplomatic documents.
   d. Trying to find ways to get others to compromise so we could reach a settlement.

Rubrics

A final aspect to consider when examining assessment techniques and measurement tools is the construction of an effective rubric. A well–constructed rubric provides a detailed analysis of the task to be completed, which directs students in their effort to successfully complete the assessment as well as helps the instructor to consistently and fairly mark the assessment. Therefore, it is worthwhile to examine the essential components of an effective rubric.

A rubric describes various levels of achievement along a continuum. A commonly used type of rubric is the scoring or analytic rubric that is designed with explicit descriptions of performance characteristics, which correspond to a point on a rating scale (Allen 2004; Deardorff et al. 2009). An advantage to using an analytic rubric is that students have a detailed explanation of the expectations for the assessment, and after assessments are marked they receive specific feedback for each criterion (Mertler 2001).

The elements of a scoring rubric include criteria related to each learning objective, definitions and examples to clarify the meaning of each criterion, and descriptions of the different levels of achievement for each criterion. For each of the learning objectives, knowledge, skill, or behavior that is expected of the students, the instructor should:

   1 Decide how the achievement of each criterion will be illustrated;
   2 Describe the range of possible performances or behaviors that indicate achievement from the highest to the lowest level;
   3 Label each level with such descriptors as “below expectations,” “meets expectations,” and “exceeds expectations”;
   4 Ensure that students clearly understand how the rubric will be used to assess their performance. (Mertler 2001; Allen 2004)
An analytical rubric is designed to score individual parts of the performance or product and then sum the individual scores to obtain a total score (Nitko 2001; Moskal 2000). Figure 3 is an example of an analytical–scoring rubric. Based on a critical-thinking rubric from the Washington State University Critical Thinking Project, it was developed to assess a general education outcome at Mid–South Community College in West Memphis, Arkansas (Peirce 2006). Three criteria are shown on the rubric in Figure 3 as a sample.

<table>
<thead>
<tr>
<th>Rubric component</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Sub-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies and summarizes the problem/question at issue.</td>
<td>Accurately identifies the problem/question and provides a well-developed summary.</td>
<td>Accurately identifies the problem/question and provides a brief summary.</td>
<td>Identifies the problem/question and provides a poor summary or identifies an inappropriate problem/question.</td>
<td>Does not identify or summarize the problem/question accurately if at all.</td>
<td></td>
</tr>
<tr>
<td>Identifies and assesses the quality of supporting data/evidence.</td>
<td>Provides a well-developed examination of the evidence and questions its accuracy, relevance, and completeness. Clearly distinguishes between fact and opinion.</td>
<td>Examines evidence and questions the quality. Distinguishes between fact and opinion.</td>
<td>Merely repeats information provided. Does not justify position or distinguish between fact and opinion.</td>
<td>Does not identify or assess the quality of supporting evidence.</td>
<td></td>
</tr>
<tr>
<td>Identifies and evaluates conclusions, implications, and consequences.</td>
<td>Accurately identifies conclusions, implications, and consequences with a well-developed explanation. Provides an objective reflection of own assertions.</td>
<td>Accurately identifies conclusions, implications, and consequences with a brief evaluative summary.</td>
<td>Does not explain, provides inaccurate information, or merely provides a list of ideas, or only discusses one area.</td>
<td>Does not identify or evaluate any conclusions, implications or consequences.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3  Adapted from a critical-thinking rubric developed at Mid–South Community College in West Memphis, Arkansas (Peirce 2006)

**Conclusion and Future Trends**

This essay has explored the literature on assessing a wide variety of active–learning techniques drawing on lessons from multiple disciplines. Even though the essay tries to be widely inclusive, the authors are aware that some bodies of literature have not been fully explored. It is quite evident that scholars of assessment and active learning are often not aware of studies that have been done in other disciplines or in other geographic areas. Scholars from different countries are asking many of the same questions (and often coming up with similar answers), but their work is narrowly circulated and their citations are mostly those of other scholars from their own country. For this reason, the culmination of knowledge is advancing at a slower rate than it might if scholars were more aware of studies done in different regions and in other disciplines.
Despite this caveat, there are several areas that are ripe for further study. One area is in robustly demonstrating the learning that occurs through the use of active-learning techniques in contrast with traditional teaching methods. Careful design of studies with a larger n and with some control variables can further contribute to this literature that is often anecdotal. A second area in which almost no publications are available is on the “portability” of active-learning exercises across cultures. Can some of the techniques that are becoming widely used within the United States be effectively adopted in education settings in other countries and vice versa? A final area in which further research is warranted is in assessing the use of newer media (internet, video content, etc.) as it is increasingly incorporated into the classroom. Although there are a growing number of studies in other disciplines regarding online learning, International Relations is behind the curve on this aspect of learning and assessment. We need to ask what assessment techniques and measurement tools can be used effectively to assess student work within the context of teaching and learning using this newer media.

References


**Online Resources**

“Active Learning” by Dee Fink. At [http://honolulu.hawaii.edu/int...evCom/guidebk/teachtip/active.htm](http://honolulu.hawaii.edu/int...evCom/guidebk/teachtip/active.htm), accessed May 2, 2011. Provides an explanation of active learning as two kinds of dialogue and presents an active–learning model that consists of three steps.

*Active Learning in Higher Education*. At [http://alh.sagepub.com/](http://alh.sagepub.com/), accessed May 2, 2011. Allows access to all issues of the journal from July 2000 to the most recent issue.


“Designing Surveys That Count” by Therese Seibert and Sherman Morrison. At [www.keene.edu/crc/forms/designingsurveythatcount.pdf](http://www.keene.edu/crc/forms/designingsurveythatcount.pdf), accessed May 2, 2011. This is a PowerPoint presentation in PDF format with tips for creating a reliable and valid survey.

Rcampus – Open Tools for Open Minds. At [www.rcampus.com/rubricshe...ls=build](http://www.rcampus.com/rubricshe...ls=build), accessed May 2, 2011. Rubric templates that make it easy to design a rubric for all subjects, from K–12 through higher education. Site registration is free.


**Acknowledgment**

Thanks to Vicki Golich for her insights in the development of this essay.

**About the Authors**

Carolyn M. Shaw is Associate Professor and Chair of the Political Science Department at Wichita State University. Her publications include *Cooperation, Conflict and Consensus in the Organization of American States* (2004), and articles in the *Journal of Human Rights*, the *Review of International Studies*, *Latin American Politics and Society*, *International Studies Perspectives*, the *Journal of Political Science Education*, and the *Journal of Conflict Studies*. She has served as the president of the Active Learning in International Affairs section (ISA) and is a member of the Political Science Education section (APSA).
Kay Gibson has taught at Wichita State University since 1998 in the areas of curriculum, instruction, and gifted education. Prior to that, she taught at the University of Southern Queensland in Australia for nine years. Kay completed her PhD in Curriculum Studies at the University of New England in Australia. Her research interests include the design of curriculum and instruction related to the development of global citizens in K–20 classrooms, and assessing higher-order thinking and creativity in active-learning curricula. She has published articles in international and national journals, chapters in edited books, and presented at numerous international, national, and state conferences about her research interests.

Cite this article


<http://www.isacompendium.com/subscriber/tocnode?id=g9781444336597_yr2011_chunk_g97814443365973_ss1-8>

Copyright

Blackwell Publishing and its licensors hold the copyright in all material held in Blackwell Reference Online. No material may be resold or published elsewhere without Blackwell Publishing’s written consent, save as authorised by a licence with Blackwell Publishing or to the extent required by the applicable law.