Authentic Assessment in Debate: An Argument for Using Ballots to Foster Talent-Development and Promote Authentic Learning

Sarah Stone Watt, *Pepperdine University*

Debate programs would be well served by participation in the culture of assessment in education. Debate is a high impact practice ideal for facilitating authentic learning. The activity has strong potential to bring about positive change in students through talent development. The ballot can serve as one productive tool for assessing talent development by providing direct evidence of student learning and advice for further improvement. Ballots provide direct evidence of a students' ongoing learning process by encouraging expert review of student performances over time. The balloting process could be enhanced to provide even stronger evidence by encouraging judges to offer more holistic feedback in the form of qualitative written comments.

As both college cost and societal demand for access to higher education continue to rise, people have taken a greater interest in what students are getting in return for their investment. Parents, employers, and lawmakers are increasingly concerned with whether students are learning the right skills and information, and whether college is making a difference in the lives of those who attend (Council of Regional Accrediting Commissions [CRAC], 2003). The seven regional accrediting agencies throughout the United States have responded to these concerns by calling on institutions to offer evidence of student learning and demonstrate that they use that evidence to make informed decisions regarding program improvement (CRAC, 2003). In order to meet these demands, institutions are looking to faculty to provide assessment data (Bresciani, 2006).

The increased emphasis on assessment has led to a redefinition of the term. Rather than simply describing a practice of student testing to evaluate learning, assessment has become the buzz word for a much larger system of holding educators publicly accountable for helping students achieve certain competencies. The language of assessment is now well engrained in institutions across the United States who recognize that everything from accreditation to financial stability may depend on their ability to demonstrate student success in meeting prescribed learning outcomes (Astin & Antonio, 2012; Bresciani, 2006). The assessment process is no longer limited to the classroom; it now reaches all corners of the institution, including co-curricular activities such as debate. As a result, debate coaches are increasingly called to justify their programs by demonstrating that debate helps students achieve core competencies or meet institutional learning outcomes.

The debate community has long touted the educational benefits of the activity. Published scholarship outlines debaters' gains in critical thinking, research skills, academic achievement, and career success (Chandler & Hobbs, 1991; Colbert & Biggers, 1985; Greenstreet, 1993; O'Donnell et al., 2010; Rogers, 2005). However, in the current era of assessment, existing broad-based studies are no longer enough evidence to demonstrate the value of debate as a learning experience. Schools and accrediting associations are looking for faculty to speak the language of assessment and, in that language, offer direct and indirect evidence that demonstrates a causal connection between individual teams' "educational practice and educational outcomes" (Astin & Antonio, 2012, p. x). While high school programs and Urban Debate Leagues have made progress in this regard (Arbenz & Beltran, 2001; Billman, 2008; National Association for Urban Debate Leagues

[NAUDL], 2013), the debate community in general lacks mechanisms for consistent assessment of its programs, and college programs in particular lack sufficient assessment data to demonstrate their role in fulfilling national and institutional learning objectives.

In this current climate, debate educators would do well to recognize their involvement in what assessment scholars call "authentic learning." Researchers have found that students learn more by doing than by merely listening and encourage educators to offer students opportunities to experiment with complex issues they are likely to face as they enter the workforce (Lombardi, 2007). Authentic learning differs from traditional learning, which is focused primarily on knowledge acquisition, in that it emphasizes the way that students employ knowledge to solve problems within particular "communities of practice" (Brown, 1999, para. 31). Authentic learning typically focuses on "complex problems and their solutions, using role-playing exercises, problem-based activities, case studies, and participation in virtual communities of practice" to prepare students for life outside of the academic environment (Lombardi, 2007, p. 2). Lombardi (2007) posits that authentic learning is particularly valuable when it enables students to make connections that cut across disciplines. Authentic learning is more likely to occur in the context of a "high-impact" practice in which students integrate diverse theories and concepts, share ideas with faculty and peers outside of class, judge the value of information, and learn from perspectives other than their own (Brownell & Swaner, 2009). These practices have been shown to help students "earn higher grades and retain, integrate, and transfer information at higher rates" (Kuh, 2009, p. 14). While it may seem obvious to debaters and coaches that the activity embodies both authentic learning and high-impact

practices, debate practitioners have not all managed to translate that message to administrators, parents, and legislators, who make important decisions regarding co-curricular programs. The way to do that is through the regular reporting of assessment data concerning debate.

At the 2012 National Communication Association Convention, Cross Examination Debate Association (CEDA) President Sarah Partlow Lefevre (2012) identified a long list of ways that debate coaches could gather information that would be beneficial for assessment. She explored the use of pre- and post-tests for topic knowledge, student evaluations, cum sheets, rankings, and more. Her talk demonstrated that there are indeed a vast number of options for coaches looking to assess their program. However, for a coach just getting started on the process, the sheer number of options may be overwhelming. Additionally, many of the available assessment tools add a significant amount of labor to an already demanding job. To make assessment both beneficial and manageable for debate teams, practitioners need an efficient tool that is adaptable to the learning outcomes of diverse programs, provides valuable evidence of student learning, and enables them to directly assess authentic learning. I argue that the ballot is an ideal assessment tool to meet these needs. In order to demonstrate the value of the ballot in assessment, I begin by reviewing some of the literature on assessment and its goals. Then, I argue that the ballot is both a means of providing programs with direct evidence of authentic student learning and a tool for program improvement because it provides expert feedback that fosters talent-development. Further, I contend that some modifications can and should be made to existing practices to make the ballot an even stronger source of assessment data.

Literature Review

Regional accrediting agencies have agreed that their goal is to "focus on the quality of student learning without specifying, beyond general categories, what that learning should be—in short, to promote *standards* without *standardization*" (CRAC, 2003, p. 1). This means that schools are still, in many ways, free to set their own goals as long as they meet them. However, completion rates alone are not evidence of meeting acceptable standards. Schools must show that they are assessing different kinds and levels of student learning. As a result, the Association of American Colleges and Universities (AAC&U) has identified "essential learning outcomes" that cut across disciplinary boundaries and "are relevant to work and life in the 21st century" (2012, p. 1). The AAC&U learning outcomes focus on: "inquiry and analysis; critical and creative thinking; integrative and reflective thinking; written and oral communication; quantitative literacy; information literacy; intercultural understanding; teamwork and problem solving" (2012, p. 1). These outcomes reflect overarching aims of higher education that ought to guide institutions in formulating their own desired learning outcomes.

Establishing institutional learning outcomes, as well as learning outcomes at lower levels (program, course, student, etc.), offers schools the chance to articulate their values by explaining the results they seek to attain. Additionally, these assessment models "focus attention on the fundamental problems of defining and measuring those outcomes that are relevant to the goals of the educational program in question" (Astin & Antonio, 2012, p. 32). In thinking about assessment of the educational experience, it is important to examine a variety of factors that influence student learning. Some factors relevant to outcome achievement are the preexisting talents a student brings with them when entering the program, the

experiences a student has in the program, the resources invested by institutions and educators into a student's experience, the program's means of facilitating student learning, and the talents a student emerges with at the culmination of their experience (Astin & Antonio, 2012; Dwyer, Millett, & Payne, 2006).

Too often, educators only measure the achievement of outcomes at the culmination of an educational experience. However, focusing only on the end result of the educational experience tends to produce data without a clear standard for discerning what the data means. For example, it is common for professors to offer a final exam to assess how much students learned from the course. Yet, Astin and Antonio (2012) explain

> All of us who have taught college students over the years know well that if a student is sufficiently bright and talented at the start of the course, it is possible for that student to do quite well on a final exam without really learning much of anything in the course. On the other hand, it is possible that a student whose performance on the final exam is mediocre may, in fact, have learned a great deal in the course, especially if the student began the course with no knowledge of the subject matter and with minimal examination performance skills. (pp. 32-33)

This example demonstrates that focusing only on a culminating measurement, such as an exam or a win/loss record, to measure student learning outcomes not only prevents educators from accurately assessing how much students are learning, but also denies educators the ability to consider environmental factors that make a particular practice succeed or fail for certain students.

An equally problematic trend in the assessment of higher education is the implicit assumption that excellence is primarily tied to financial resources and

academic reputations. This assumption is reflected in prominent college ranking systems such as those of *U.S. News and World Report* (Morse, 2012). It is true that institutions with more financial resources and higher institutional rankings may have an easier time attracting highly sought after faculty and students with high grades and test scores. However, neither of these factors demonstrates what students are learning when they enroll in the institution (Astin & Antonio, 2012). Similarly, debate programs with more resources and higher rankings may recruit talented high school debaters who are able to perform well in college-level competitions, but those programs still need to be able to demonstrate what and how their students are learning as much as a program with fewer resources and lower overall rankings that primarily focuses on novice recruitment from within the institution.

Given the prevalence of the outcomes-based model of assessment, the debate community would do well to find a way to demonstrate that the activity plays a role in training students to achieve national and institutional learning outcomes. However, the community must take care not to reproduce the mistakes made by educators in assessing outcome achievement. While debate educators may not provide final exams, measures such as win/loss records, speaker points, or placement in a given tournament or season may provide comparable forms of culminating evidence. Furthermore, debate educators may be tempted to examine the correlation between resources and reputation as measures of program excellence. While all of these variables are likely to contribute to the larger picture of what it takes to succeed in debate competition, they should not overshadow the importance of assessing what and how students are learning within the debate context.[1]

Rather than understanding program excellence

based primarily in competitive success, students and institutions would be better served if debate educators increased their focus on the authentic learning process and defined a program's excellence based on the "ability to bring about positive change in students," which "should be measured in terms of the growth and improvements in students over time" (Tam, 2002, p. 214). This model, which Astin (1993) calls the talent-development conception of excellence, measures program excellence by the ability to develop talent regardless of a students' starting place. This model would allow debate educators to combine existing measurements of success with an emphasis on sustained feedback and evaluation over time to see that students experience authentic learning.

Students come to debate with diverse educational backgrounds and varying levels of experience with the activity itself. Debate educators would be better able to assess their program excellence, translate their findings to administrators, and make strategic decisions regarding program improvement if the community focused on a talent-development approach. The talent-development approach relies on the ability to demonstrate that debate has an impact on "students' knowledge and personal development" (Astin & Antonio, 2012, p. 7). To demonstrate this, programs need to focus on the various ways that debaters change and improve over time.

The assessment process can contribute to talent development by providing direct feedback to students and indirect feedback to coaches regarding "the effectiveness of various educational practices" (Astin & Antonio, 2012, p. 252). Astin and Antonio (2012) contend that these goals are best achieved through a performing arts theory of education wherein students and educators receive direct feedback concerning their performance and advice on ways to improve. This

feedback should be based on "holistic judging," in which student performance is judged based on a number of different dimensions. Unfortunately, "one problem with holistic written evaluations of student performance . . . is that they do not readily yield quantitative estimates of student performance" (Astin & Antonio, 2012, p. 60). Some people also object to holistic judging because they argue that the process is too labor-intensive. In the remainder of this essay, I will characterize the ballot as an existing source of holistic judging that, if used properly, can overcome these criticisms and enable debate educators to focus on a talent-development model of assessment that not only demonstrates program excellence, but also positions them to articulate their role in fostering an authentic learning environment that contributes to the fulfillment of both national and institutional learning outcomes.

The Role of the Ballot

More than two decades of research has established the existence of certain "high-impact practices" which demonstrate "the value of active, engaged, and collaborative forms of learning for students" (Kuh, 2009, p. 32). High-impact practices are characterized by a number of elements which are central to academic debate. High-impact programs tend to:

Demand that students devote considerable time and effort to purposeful tasks; most require daily decisions that deepen students' investment in the activity as well as their commitment to the academic program and college. . .demand that [students] interact with faculty and peers about substantive matters, typically over extended periods of time. . . participating in one of these [high impact] activities increases the likelihood that students will experience diversity through contact with people

> who are different from themselves. . . even though structures and settings of high impact activities differ, students typically get frequent feedback about their performance in every one. . . participation in these activities provides opportunities for students to see how what they are learning works in different settings, on and off campus. Finally, it can be life changing. (Kuh, 2009, p. 17)

These high impact practices are ideal for facilitating authentic learning. The current list of high impact practices includes activities like undergraduate research, first year seminars, learning communities, collaborative assignments and projects, and diversity/global learning, but fails to account for academic debate. George Kuh (2011), founding director of the National Survey for Student Engagement, recognizes that co-curricular activities such as debate may be high-impact practices, but they were not included in the AAC&U report on high-impact educational practices because "we don't have research. . . that documents their impact" (personal communication, October 22, 2011). Waters and Stone Watt (2012) explain that there is a need for debate programs to establish and measure learning outcomes that not only demonstrate authentic learning, but that also showcase the activity's ability to draw students from diverse majors, at different points in their educational experience, and with differing levels of activity-specific knowledge together in a collaborative learning environment.

Debate practitioners need assessment data to demonstrate that the activity is a "high-impact" practice. The debate ballot provides an opportunity to collect data that demonstrates authentic learning focused on communal problem solving, one that can be adapted to speak to institutional and program learning outcomes, and can provide holistic judgment and direct feedback

to foster "talent development." The ballot is an ideal starting place for assessment because it provides direct evidence of student learning through regular evaluation by external expert reviewers at various stages of student development throughout a given school year.

Direct Evidence

At its core, assessment is supposed to provide educators with evidence of what is working and what needs improvement. Astin and Antonio (2012) explain that, "assessment results are of most value when they shed light on the causal connections between educational practice and educational outcomes" (p. x). Evidence of these causal connections tends to come in two forms: indirect and direct. Indirect evidence is based on students' reflections on learning. It calls for students to interpret how their knowledge and skills have been enhanced by the learning experience. Direct evidence comes from "analysis of student behaviors or products" (Allen, 2008, p. 1), in which students are judged based on the knowledge and skills they demonstrate. This type of evidence can offer quantitative scores or rankings, as well as provide qualitative description regarding the interpretation of student knowledge and behaviors (Palomba & Banta, 1999). The ballot is designed to provide direct evidence of student learning.

Ballots offer three quantitative measures of student learning: (1) win/loss, (2) speaker rankings, and (3) speaker points. The debate community is adept at evaluating teams based on their wins and losses at various tournaments. Coaches, who generally have experience as both debaters and judges, are able to evaluate their own teams and benchmark against other teams based not only on win-loss records, but also the quality of a win or loss based on the quality of the tournament, opponent, and critic for a given debate.

They are also able to assess improvements in student performance based on numerical speaker rankings and points. While speaker rankings are easy to evaluate, points are more ambiguous. Speaker points were once calculated based on a rubric provided at the top of the ballot (Cirlin, 1986; Leeper et al., 2010). This system did not completely prevent ambiguity in the assignment of points, but it did make them more meaningful than they are today. Currently, speaker points may be given on a 30 or 100 point scale, depending on the tournament, and there is very little consensus regarding what a particular score means. Judges are free to evaluate speaker points differently based on personal opinion, level of competition, national versus local tournaments, judge or debater identity, and more (see, for example, Galloway, 2010; Kelsey, Evans, Marty, & Reid-Brinkley, 2012; Rubaie, 2011). Leeper et al. (2010) conclude that speaker points today "have become disconnected from meaningful criteria, and, instead, reflect a general sense of whether the debater is competent" (p. 152). Since the standards for competence in this model are debatable, it is difficult to make causal connections between speaker point fluctuations and student learning. The lack of consensus on this issue heightens the need for additional qualitative feedback to help with ballot assessment.

Until fairly recently, judges were obligated to take time filling out the bottom of the debate ballot with reasons for their decision. They provided written feedback concerning both the arguments and students' performance in the debate. Unfortunately, "the written ballot as a tool to convey the reason for decision has gone the way of the dodo" (Leeper et al., 2010, p. 152). Judges now tend to offer verbal feedback after the round, which is valuable in giving students an immediate reason for their win or loss and helping them identify improvements they might make as the tournament goes on. While there

are clear benefits to this model of post-round discussion, there are two apparent drawbacks that are pertinent to assessment: (1) the amount of information can be overwhelming for students to process, particularly on the heels of an intense two hours of debate; and (2) coaches are rarely able to sit and listen to or record each of their teams' post round discussions. As a result, students are free to take the advice that resonates with them, and the information that makes it back to the coach is, if anything, only partially reflective of the judge's feedback. While the debate community would "never hope to lose the in-depth critical review of a student's arguments that a conversation about a specific debate can provide" (Leeper et al., 2010, p. 152), a return to qualitative feedback on the ballot would provide a much needed supplement to these conversations and offer programs the direct evidence they need to draw causal connections between their educational practices and the outcome of a given debate. The increasing use of electronic balloting on platforms such as www.tabroom.com should make it easier for judges to document their feedback because they can type directly into the ballot during the debate and edit their comments before saving them (Bruschke & Nielson, 2006). These platforms also have the potential to provide both coaches and national organizations with a database of the quantitative and qualitative data for both assessment and broader research purposes.[2]

Expert Review of Authentic Learning

The ballot not only provides direct evidence of student learning, it offers a space for debate experts to give feedback on authentic learning. Schools have long focused on high impact practices such as internships and apprenticeships to evaluate authentic learning, and have recently begun to include other performance based activities in that category (University of Central Florida, 2012). Despite its emphasis on complex

simulations of real world problem solving, debate is one performance based activity that is generally ignored in these discussions. The qualitative comments on a debate ballot could offer programs the evidence they need to demonstrate that debate is an authentic learning experience by providing evidence of the multiple, varied, and complex problem-solving tasks students are engaged in throughout the course of each tournament. These comments are arguably more valuable than the average authentic assessment measure because they come from well qualified external reviewers.

While most authentic assessment is done by faculty members evaluating their students' performance, debate teams have the unique advantage of supplementing their own assessment with multiple expert reviews each time they compete. Whereas authentic assessment of activities such as internships may benefit from a single external expert review of student performance in the form of a supervisor evaluation at the culmination of the internship, debaters are evaluated by at least six different judges with expertise in academic debate at every tournament they attend. The average college debate tournament judging pool consists of former debaters, graduate assistants, and faculty coaches, most of whom have (or are in the process of obtaining) advanced degrees in their fields, and have multiple years of experience in the activity. While individual teams' values influence their perception of judge expertise, that vast majority of debate judges meet the standard for "expert review" in that they possess specialized knowledge of the activity and the standards by which it ought to be evaluated (Case, Jorgensen, & Zucker, 2004). Judges bring diverse knowledge and perspectives to the assessment process. Rather than simply viewing the post-round critique as advice on how to win the next debate, debate judges ought to also be seen as providing authentic

assessment data to be used in program improvement.

In order to assess authentic learning experiences, Herrington and Herrington (2008) explain that "the learning environment needs to ensure that the assessment is seamlessly integrated with the activity and to provide the opportunity for students to be effective performers with acquired knowledge, and to craft products or performances in collaboration with others" (p. 73). The debate ballot provides a space where the judge can offer an evaluation of a debater's knowledge, performance, collaboration with their partner, and/ or adaptation to the opposing teams' arguments. While the judge's decision in a round is final, the ballot is a reference point for collaboration whereby debaters, coaches, and judges can work together to refine a particular argument or performance. Beyond providing input on a student's choice of argument, judge feedback also tends to offer advice on additional areas for research, alternative perspectives on an issue, and community norms. This begins an iterative process whereby the debater learns to adapt not only to a particular judge, whom they will likely encounter again, but to a given perspective on debate. Rather than being a one-time assessment at the end of an experience, each debate ballot is a small piece of the ongoing authentic learning and assessment process wherein the student presents what they have learned, solves problems by responding to opponents, is evaluated by an expert, and uses that feedback in collaboration with teammates and coaches to add to their knowledge and/or modify their performance in anticipation of the next evaluation.

Qualitative feedback on ballots could aid coaches with two important elements of the authentic assessment process: scaffolding and reporting. Scaffolding in the coaching process is a crucial part of authentic learning because it calls on educators to avoid a didactic approach

in favor of one that "provides for coaching at critical times, and scaffolding of support, where the teacher and/or student peer mentors provide the skills, strategies, and links that the students are unable to provide to complete the task" (Herrington & Herrington, 2008, p. 73). While debate coaches may already engage in scaffolding to help students improve, the written ballot provides a means of formalizing that process in a way that better lends itself to institutional reporting. Currently, a coach may scaffold by hearing a post round critique or talking to a debater about what they are struggling with and then determining the best coach or peer on the team to mentor the debater in their attempt to resolve the problem. While this process may occur at multiple points throughout a tournament or team practice session, there is rarely an opportunity for the coach to reflect on and record the process. By collecting written ballot comments over the course of a season, coaches can compile written evidence of student learning and areas for improvement. They can not only use this evidence to improve their coaching and scaffolding throughout the season, but also to report to their institution what expert reviewers have said regarding their successes and areas for improvement. At the end of each debate season, coaches should be able to use excerpts from qualitative ballot comments as evidence to demonstrate the authentic learning process and fulfillment of stated learning outcomes to their institution.

Holistic Judging and Talent-Development

Assessment scholars have long criticized the testing model "for providing inadequate or superficial feedback to students" (Lombardi, 2008, p. 4). When debaters focus solely on the win or loss in a given debate, we reproduce the problem. Instead, Leeper et al. (2010) argue that it may be time to move beyond a model of competition so focused on winning and begin to ask ourselves what

practices we want to incentivize. The debate community seems to be doing this, to some extent, in post round discussions and deliberation on CEDA Forums concerning the implications of arguments and judging criteria on participation and success in the activity. Increased reliance on the ballot, and qualitative written feedback, as a tool of assessment offers teams the ability to understand wins and losses in an effort to improve talent development.

The ballot offers a site where judges can express their holistic evaluation by providing feedback on multiple facets of the debate that may not be possible in a short post-round discussion. The current model of post-round discussion assists in talent-development by offering "direct feedback to the learner" (Astin & Antonio, 2012, p. 252), but it does not go far enough in providing indirect feedback to the coach. The typical post-round discussion is a condensed time in which debaters and judges are still processing the outcome of the debate, packing up their belongings, and thinking of what comes next, whether it is a meal or another debate. In this short time, the judge often attempts to share not only their rationale for deciding who won and lost, but also other pieces of advice about arguments, style, research, and more, as well as address debater questions and document any arguments or source citations they might want for their own coaching purposes—this is a tall order. This approach fails to recognize that in learning, sometimes less is more. Researchers have found that educators attempting to cover "vast amounts of material may actually *impede* understanding and lower student achievement" (Lombardi, 2008, p. 4). This may be why, after a detailed post round discussion, debaters report back to coaches that they lost the debate because the judge didn't understand their arguments, or the judge was biased, without reflecting the thoughtful advice

they were just given. Writing feedback on the ballot, rather than trying to discuss everything with students following the debate, would allow judges to focus their post-round critique on smaller amounts of information that are immediately useful to debaters in a given competition, and easy to digest. They could leave more complex feedback on the ballot for debaters and coaches to process later, when they are working on improvements for the next tournament.

Debate ballots already address the primary critique of current holistic evaluations of student performance by providing quantitative feedback. Qualitative feedback on the ballot would assist coaches with the second portion of the talent development model by "informing the practitioner about the effectiveness of various educational practices" (Astin & Antonio, 2012, p. 252). The debate community is intimate enough that judges are occasionally able to speak directly to a particular coach's pedagogical practice, but it would be unreasonable to expect every judge to do so in every debate. Instead, qualitative comments highlight important moments in a given debate, which can help coaches reading the ballot later to determine whether their team's performance in the round reflects an understanding of the coaching they have received. Qualitative feedback on the ballot may help educators to draw the necessary connections between outcomes such as win/loss records, rankings, and speaker points, and the debaters' arguments, and performance in the debates themselves. It would also provide a record, over time, that coaches could analyze to establish recurring factors evident in a students' performance that point to success or failure in meeting established learning outcomes.

Limitations and Suggestions for Future Research

The primary criticism of the holistic judging model

is that it is too labor intensive (Astin & Antonio, 2012). However, debate judges are already engaged in this labor intensive process; the ballot simply provides a means of recording their existing efforts. The recent increase in online balloting also mitigates this concern because it allows the judge to type comments into a form as the debate progresses and edit them before sending.

Shifting ballots to an online platform provides opportunities to overcome the limitations of the existing ballot format for research purposes. However, in addition to the need for judges to write more qualitative feedback on their ballots, there are two modifications that would make the ballot an even more useful tool for assessment. First, the debate community should adopt a universal speaker point scale. Multiple debates over what particular points on a given scale mean to individual judges have demonstrated our inability to reach consensus about the difference between, for example, a 27.5 on the 30 point scale and 70 on the 100 point scale. Since codifying a particular interpretation of the numbers on a given scale, or equating numbers on one scale to numbers on another, has proven to be a losing battle, we would do well to at least pick a scale and stick with it so that we are comparing a similar method of scoring across all tournaments in a competitive season. Second, to counterbalance the ambiguity of speaker points, we should also return to some form of rubric embedded in the ballot. Astin and Antonio (2012) suggest that a simple way of making quantitative and qualitative measures more useful is to "have the evaluator also complete a brief set of rating scales, with each scale representing a different skill, area of knowledge, or personal quality" (p. 60). They liken this approach to scoring essays or adjudicating musical and artistic competitions. This is an approach seasoned debate judges are familiar with from the

American Forensics Association (AFA) ballots, "a long standard in the debate community" (Cirlin, 1986, p. 86). The AFA categories such as delivery, analysis, reasoning, and evidence, are a good start, but the advancements in rubric creation and electronic balloting mean that we could develop something that better serves the community's assessment needs. The AAC&U's essential learning outcomes may provide a good starting place for the development of a rubric with clear and concise evaluation criteria and a simple Likert scale for evaluation. To improve the usefulness of the rubric, the online ballot could either display or link to a set of quality definitions which provide a more detailed explanation of what a debater must demonstrate in order to achieve a particular placement on the scale (Reddy & Andrade, 2010).

The CEDA Research and Assessment Working Group (2013) is examining two important changes in this regard. First, they are drafting prospective learning outcomes for the activity that address some of the goals outlined by AAC&U and reflect some of the unique skills debaters acquire. Second, they are considering how these learning outcomes might be measured through the use of a rubric embedded in the online ballot whereby judges can offer brief but pointed feedback concerning the particular performance they observed in a given round. Both the general learning outcomes and rubric process still require additional research and testing before they can be implemented to improve on the existing ballot system.

One additional issue that must be addressed as the debate community continues to consider the best ways to assess ballots and other forms of evidence demonstrating student learning is the need for informed consent. Typically, if assessment data is collected by an educator "to provide feedback to students, improve a course or

program, or provide findings to . . . [an institutional assessment office] for university-wide educational program improvement" (Virginia Tech Institutional Review Board, 2013, para. 1), they do not need approval from an institutional review board (IRB). This means that coaches should be able to assess their teams' ballots for the purpose of improving their program and reporting program data to their own institution. However, if the debate community wishes to make ballots publicly available for wider assessment purposes, or utilize data from ballots in published research or grant applications, researchers may need to seek IRB approval and devise a means of obtaining informed consent before analyzing ballots.

Conclusion

After decades of development, assessment is here to stay, and is increasingly being used to hold institutions and programs accountable (Weiner, 2009). Debate programs are in need of evidence of student learning that demonstrates their ability to meet established learning outcomes. Furthermore, programs would be well served by assessment mechanisms that not only address learning outcomes, but position debate as a co-curricular activity uniquely suited to meet increasing demands for high-impact authentic learning experiences. The ballot has the potential to meet these needs if it is used strategically.

The ballot has the potential to serve as an ideal assessment measure for multiple reasons. First, it offers direct evidence of student learning in the form of both quantitative and qualitative feedback that can address varied learning outcomes. Second, it provides multiple opportunities for external expert review of student learning. Third, it refocuses assessment on talent-development rather than resources and reputation.

The ballot assists in talent-development by offering direct feedback to the student while also conveying that information to the coach for use in improving their educational practices. Fourth, the collection of ballots over time provides evidence of students' knowledge and skills development across a season, or even a full debate career. Fifth, the ballot offers debate advocates a space from which to demonstrate that their students are engaged in high-impact and authentic learning. While each debate team will find it necessary to supplement this assessment measure with appropriate forms of indirect evidence that they are meeting or exceeding institutional goals, the ballot is an existing measure that gives every team direct evidence of student learning.

With an increased commitment to providing detailed ballot feedback, the debate community could offer programs the ability to document their success in meeting national and institutional learning outcomes. In order to begin using ballots in this way, debate programs should first document their goals in the language of assessment by crafting a written set of student learning outcomes. The program's student learning outcomes should align with those of the institution and be stated in such a way that they,

> Specify what students will know or be able to do as a result of the activity. . . They should be observable (and when appropriate measureable). They should be clearly written and easily understood by faculty, students, staff, administrators, and even parents. (Office of Institutional Effectiveness [OIE], 2013, para. 2)

The statements should employ verbs that describe what actions students should demonstrate by the end of a given season, or their participation in the activity (OIE, 2013). Throughout the season, coaches should

make note of the ways that students demonstrate their achievement of, or need for assistance in achieving, the stated learning outcomes. This is where ballot comments and, perhaps in the future, rubrics will be useful as evidence for the coach's observations. They should also document the strategies and scaffolding used to assist students in achieving the outcomes. At the end of the season, coaches should draft a report for their institution, using ballot data as one form of evidence, which indicates whether and how their program is achieving the student learning outcomes set at the beginning of the season. This is also an opportunity for coaches to be honest about where their program may fall short of meeting program or institutional objectives and consider what improvements are necessary.[3]

Ultimately, assessment should provide feedback to an educator that helps them gauge the "impact of their educational practices and policies" and make adjustments to enhance their effectiveness (Astin & Antonio, 2012, p. 141). The ballot is an existing tool designed to do exactly that. However, debate practitioners are not currently using the ballot to their strategic advantage in assessment. With some minor modifications, the debate ballot could provide not only a way to assess and improve individual programs, but also to make larger scale arguments concerning the benefits of debate across educational contexts.

Notes

[1] Research in the field of education suggests that students do show higher improvement in performance when attending esteemed, "resource-rich" institutions (Arum & Roska, 2011; Astin, 1993; Astin & Antonio, 2012).

[2] As of this writing, judges can type comments into ballots on tabroom.com, but most don't. Those comments that are made are not currently available to coaches, but I understand that this is an issue administrators are addressing.

[3] Improvements may be things that the program itself needs to work on, or they may indicate a need for assistance from the institution. For example, our program successfully relied on assessment data to demonstrate the need for resources to hire an additional coach.

References

Allen, M. J. (2008, July). *Strategies for direct and indirect assessment of student learning*. Paper presented at the Southern Association of Colleges and Schools Commission on Colleges in Orlando, FL. Retrieved from: http://assessment.aas.duke.edu/documents/ DirectandIndirectAssessmentMethods.pdf

Arbenz, C., & Beltran, S. (2001). *Empowering Latinas through debate: An analysis of rates of success at SCUDL tournaments of Latina debaters*. Paper presented at the National Communication Association Convention in Atlanta, GA. Retrieved from: http:// communications.fullerton.edu/clubs/forensics/ papers.asp

Arum, R., & Roska, J. (2011). *Academically adrift: Limited*

learning on college campuses. Chicago: University of Chicago Press.

Association of American Colleges and Universities (2012, February). *A sea change on student learning assessment: An AAC&U working paper.* Retrieved from: http://www.aacu. org/resources/assessment/documents/ AACUAssessmentConceptPaper5-18-2012.pdf

Astin, A.W. (1993). *What matters in college? Four critical years revisited.* San Francisco, CA: Jossey-Bass.

Astin, A.W., & Antonio, A. L. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education,* 2nd ed. Lanham, MD: Rowman & Littlefield Publishers.

Billman, J. (2008). Your community doesn't have to win nationals: The benefits of forensics to parents, communities, and society. *Rostrum, 82*, 97-99.

Bresciani, M. J. (2006). *Outcomes-based academic and co-curricular program review.* Sterling, VA: Stylus Publishing.

Brown, J. S. (1999). *Learning, working, and playing in the digital age.* Talk presented at the Conference on Higher Education of the American Association for Higher Education. Retrieved from: http://serendip. brynmawr.edu/sci_edu/seelybrown/seelybrown.html

Brownell, J. E., & Swaner, L. E. (2009). High-impact practices: Applying the learning outcomes literature to the development of successful campus programs. *Peer Review, 11,* 26-30.

Bruschke, J., & Nielson, T. (2006). Online debate record keeping and judge information: An imprudent look at the future. *Contemporary Argumentation and Debate, 27,* 121-132.

Case, B. J., Jorgensen, M. A., & Zucker, S. (2004). Alignment in educational assessment. *Pearson Assessment Report*. Retrieved from: http://www.pearsonassessments.com/NR/rdonlyres/C19A3D92-B124-4098-BC2F-0BCD490DCAA2/0/AlignEdAss.pdf

Chandler, R.C., & Hobbs, J. D. (1991). The benefits of intercollegiate policy debate training to various professions. In D. Parson (Ed.), *Argument in controversy: Proceedings of the Seventh SCA/ AFA Conference on Argumentation* (pp. 388-390). Annandale, VA: Speech Communication Association.

Cirlin, A. (1986). Judging, evaluation, and quality of CEDA debate. *The National Forensic Journal, 4,* 81-90. Retrieved from: http://www.nationalforensics.org/ journal/vol4no2-1.pdf

Colbert, K., & Biggers, T. (1985). Why should we support debate? *Journal of the American Forensics Association, 21*, 237-240.

Council of Regional Accrediting Commissions (2003, May). *Regional accreditation and student learning: Principles for good practices*. Washington, DC: CRAC. Retrieved from: http://www.msche.org/ publications/Regnlsl050208135331.pdf

Cross Examination Debate Association (CEDA) Research and Assessment Working Group (2013, November). *Report on the findings of the CEDA Research and Assessment Working Group*. Presented at the annual meeting of the National Communication Association Convention, Washington DC.

Dwyer, C. A., Millet, C. M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes*. Princeton, NJ: ETS. Retrieved from: http://www.csus.edu/programassessment/Assessment%20Plan%20Rubric%20Q%27s/ContextPieces/

Political%20Context/cultureofevidence.pdf

Galloway, R. (2010). Speaker points, packets, and ballots: The lost year [Discussion Forum]. Retrieved from: http://www.cedadebate.org/forum/index. php?topic=460.0

Greenstreet, R. (1993). Academic debate and critical thinking: A look at the evidence. *National Forensic Journal, 11*, 13-28.

Herrington, A., & Herrington, J. (2008). What is an authentic learning environment? In T. Herrington & J. Herrington (Eds.), *Authentic learning environments in higher education* (pp. 68-77). Hershey, PA: Information Science Publishing.

Kelsey, A., Evans, R., Marty, J., & Reid-Brinkley, S. (2012, November 12). *An open letter to Sarah Spring*. Retrieved from: http://resistanceanddebate. wordpress.com/2012/11/12/an-open-letter-to-sarah-spring/

Kuh, G. D. (2009). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington DC: Association of American Colleges and Universities. Retrieved from: http:// www.neasc.org/downloads/aacu_high_impact_2008_ final.pdf

Leeper, K., Abbott, B., Congdon, K., Gonzales, J., Partlow Lefevre, S., Richardson, D., ...Varda, S. (2010). Innovation and debate: Where do we go from here? In A. D. Louden (Ed.), *Navigating opportunity: Policy debate in the 21st century* (pp. 149-162). New York: International Debate Education Association.

Lombardi, M. M. (2007). *Authentic learning for the 21st century: An overview*. Educase Learning Initiative. Retrieved from: http://net.educause.edu/ir/library/

pdf/eli3009.pdf

Lombardi, M. M. (2008). *Making the grade: The role of assessment in authentic learning.* Educase Learning Initiative. Retrieved from: http://net.educause.edu/ir/library/pdf/eli3019.pdf

Morse, R. (2012, September 11). Methodology: Undergraduate ranking criteria and weights. *U.S. News and World Report.* Retrieved from: http://www.usnews.com/education/best-colleges/articles/2012/09/11/methodology-undergraduate-ranking-criteria-and-weights-2

National Association for Urban Debate Leagues (2013). *The 2013 Urban Debate National Championship and Annual Dinner: A Report to Friends and Funders of the National Association for Urban Debate Leagues.* Retrieved from: http://www.urbandebate.org/pdf/NAUDL%202013%20Tournament%20Report.pdf

O'Donnell, T., Butt, N., Bauschard, S., Bellon, J., Decker, W., Katsulas, J., Keith, W., ...Packer, J. (2010). A rationale for intercollegiate debate in the twenty-first century. In A. Louden (Ed.), *Navigating opportunity: Policy debate in the 21ˢᵗ century* (pp. 27-56). New York: International Debate Education Association.

Office of Institutional Effectiveness. (2013). *Writing SLOs.* Pepperdine University. Retrieved from: http://services.pepperdine.edu/oie/learning-outcomes/writing-slos.aspx

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education.* San Francisco, CA: Jossey-Bass Publishers.

Partlow Lefevre, S. (2012, November). *Documenting debate: Information gathering and record keeping for*

assessment of intercollegiate debate programs. Paper presented at the annual meeting of the National Communication Association, Orlando, FL.

Reddy, Y. M., & Andrade, H. (2012). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education, 35*, 435-448.

Rogers, J. E. (2005). Graduate school, professional, and life choices: An outcome assessment confirmation study measuring positive student outcomes beyond student experiences for participants in competitive intercollegiate forensics. *Contemporary Argumentation and Debate, 26*, 13-40.

Rubaie, B. (2011, November 17). Speaker points— Fall 2011 [Online discussion forum]. Retrieved from: http://www.cedadebate.org/forum/index. php?topic=2978.0

Tam, M. (2002). University impact on student growth: A quality measure? *Journal of Higher Education Policy and Management, 24*, 211-218.

University of Central Florida (2012). *Authentic assessment and rubrics*. Retrieved from: http://www.fctl.ucf. edu/teachingandlearningresources/coursedesign/ assessment/assessmenttoolsresources/rubrics.php

Virginia Tech Institutional Review Board. (2013). *FAQs for academic assessment research*. Retrieved from: http://www.irb.vt.edu/pages/assessment.htm

Waters, K., & Stone Watt, S. (2012) *Assessing high-impact co-curricular programs in the communication disciplines: Creating consistent SLOs, overcoming challenges, and planning for the future.* Presentation at the Academic Resource Conference for the Western Association of Schools and Colleges, Costa Mesa, CA.

Weiner, W. F. (2009). Establishing a culture of assessment.

Academe, 95, 28-32.

Sarah Stone Watt (Ph.D., The Pennsylvania State University) is Director of Forensics and assistant professor of communication at Pepperdine University. Correspondence should be addressed to Sarah Stone Watt at Sarah.StoneWatt@pepperdine.edu.