

How Does Applied Learning Fare Under the Scrutiny of Program Review?

EMILY J. KEMPF

Rockhurst University

ROY SCHWARTZMAN

University of North Carolina at Greensboro

ERIN WILSON

KELLY BOUAS HENRY

Missouri Western State University

An educational environment characterized by shrinking fiscal and physical resources has spurred many institutions to undertake comprehensive academic program review. Such a process assesses the performance of programs and prioritizes them for de-emphasis, maintenance, or enhancement. How can applied learning be properly acknowledged within program review? On the macro level, performance criteria can be re-envisioned to reward the pedagogical practices embodied in applied learning. On the micro level, an applied learning assessment instrument is discussed and tested that can provide multidimensional performance data across academic fields.

The ongoing erosion of financial and other resources available to sustain higher education has led to more voluble calls for accountability, especially for public institutions that must prove themselves responsible trustees of ever-scarcer government funds (Michael, 1998). How can an institution responsibly cope with this resource scarcity? The simple, mindless answer is the one corporate America often selects: obliterate the sources of the costs. The most straightforward—and

shortsighted—method would be to impose horizontal cuts that slice a uniform proportion of expenditures across the board. This clever-like approach to expenditure reduction, while simple, is not only imprudent but fails to address systemic misallocations of resources. Horizontal cuts unduly penalize areas that might be centers of excellence while not slicing deeply enough into chronically under-performing or wasteful areas. Even worse, these uniform cuts sidestep the opportunity that lurks beneath financial crisis: the prospect for rigorous self-examination that can reorient an institution toward strategically investing in its future.

Although academic program review has borne many names and is implemented differently on each campus, it can provide a useful alternative to indiscriminate, across-the-board budgetary butchery. If responsibly developed and implemented, academic program review generates an iterative process of self-assessment, clearer focus on goals, and adaptability to change. In all its manifestations, academic program review maintains certain core characteristics. Although their methods may differ, all such reviews emphasize the need to allocate resources in ways that can be justified rationally. The basis for these justifications lies in information gathered about each academic program.

Program review has a retrospective and a prospective function. Retrospectively, it provides information about past performance of academic programs—tangible evidence of historic trends. Prospectively, program review enables an institution to reallocate its resources—all dimensions of resources—in justifiable ways according to the institution's priorities (Dickeson, 2010). Considered from the standpoint of expenditures, program review provides the functional equivalent of research that informs how an institution decides to invest in itself.

This essay consists of two distinct but intertwined components. The macro-level discussion considers how applied learning can be systematically incorporated into academic program review. As subsequent analysis will demonstrate, applied learning has several characteristics that render such pedagogy vulnerable when assessed by many of the

conventional criteria employed in academic program reviews. The macro-level issues primarily concern the systematic ways applied learning can be prioritized or penalized in the program review process. The micro-level discussion shifts attention to an alternative assessment methodology for applied learning, specifically an internally developed instrument for measuring quality that focuses on features indigenous to experiential pedagogies. The micro-level component focuses on the tools that can generate evidence for how well (or poorly) applied learning is performing, furnishing the basis for assessments of quality that can drive concomitant resource allocation. Finally, the macro and micro levels are synthesized in reflections about the compatibility between applied learning and academic program review.

MACRO-LEVEL: ALIGNING APPLIED LEARNING WITH THE ACADEMIC PROGRAM REVIEW PROCESS

Academic program review is the functional equivalent of a performance appraisal of educational delivery. This performance appraisal generates recommendations for action, and the systematic application of program review is designed to shape policy regarding expansion, consolidation, alteration, or attenuation of educational services. What position might applied learning occupy within the evaluative framework of a comprehensive academic program review? If left unexamined, academic program review could introduce systematic under-acknowledgment of applied learning pedagogies. This marginalization, while unintentional, might leave unnoticed many of the connections between applied learning and fundamental life skills (Schwartzman, 2010).

CHALLENGES FOR APPLIED LEARNING

Applied learning could suffer under academic program review if such a review defines its unit of analysis (a "program") too narrowly and inflexibly. Applied learning as a pedagogy is transdisciplinary by nature. Of course, particular academic disciplines at a given campus may employ applied learning more than other fields. Administrative units such as an International Programs Center, Service-Learning Office, Undergraduate Research Office, or similarly designated entities may serve as focal points of applied learning practices. An institution-wide, systematic emphasis on applied learning, however, requires more than "a thousand points of light" that illuminate scattered pockets of experiential education. If the unit of analysis employed in academic program review is an academic department, major, or similar range of program delivery, the full measure and impact of applied learning could escape notice. To occupy a central place in academic program review, applied

AUTHOR NOTE: Emily J. Kempf, M.S., Department of Residence Life, Rockhurst University. Roy Schwartzman, Ph.D., Chancellor's Resident Fellow, Lloyd International Honors College, and professor, Communication Studies Department, 109 Ferguson Building, University of North Carolina at Greensboro, Greensboro, NC 27402-6170. E-mail: docroy@triad.rr.com. Erin Wilson, Missouri Western State University. Kelly Bouas Henry, Ph.D., Professor of Psychology, 217H Murphy Hall, Missouri Western State University, St. Joseph, MO 64507. E-mail: henry@missouriwestern.edu. This research was supported in part by a Strategic Planning Implementation Funds grant from Missouri Western State University.

learning needs to be articulated as something deliberately embraced and systematically implemented throughout the institution. In this role, applied learning qualifies as integral to delivering desired educational outcomes rather than merely a co-curricular add-on incidental to education. Indeed, if approached primarily as a co-curricular “enhancement,” applied learning appears as an easily severed appendage to more restrictive, classroom-bound pedagogies that fail to engage a broad range of students’ intellectual capabilities (Slavkin, 2004).

How can academic program review capture the widest range of applied learning practices? If program review remains focused within the administrative or disciplinary silos of individual fields, departments, or offices, then it becomes difficult to aggregate applied learning *per se* because these pedagogical techniques do not always align neatly with the boundaries of majors, faculty lines, or even the definitions of traditional scholarship and creative activity. To remedy this potential for applied learning to fall through the cracks of program review, an institution needs to embrace applied learning as a central priority embodied in the value-saturated mission, vision, or strategic plan that forms the philosophical basis for program review itself. Any academic program review can aggregate and evaluate applied learning practices from the top by connecting the values that guide program review to the criteria for determining the quality of academic programs.

Explicitly connecting an institution’s core values or objectives with applied learning could proceed by aligning these aspirations with applied learning practices that enable their achievement. Table 1 offers a hypothetical example of how a college or university might identify applied learning as a principal mode of achieving desired institutional goals.

It may be difficult or impossible to generate data that isolates experiential learning as the sole independent variable responsible for academic success. This challenge, however, does not automatically spell doom for applied learning. Rather, evaluators of academic programs should ask how applied learning contributes to the factors involved in achieving academic success as defined by the institution. Focus would shift away from capturing the elusive essence of quality and instead concentrate “on processes believed to produce quality” (Wergin, 2002, p. 248). Con-

sidered in more social scientific terms, perhaps applied learning needs to be treated more as a mediating factor that contributes to performance instead of as an independent cause of specific educational outcomes. Approached from this angle, program review could concentrate on how the institution’s achievement of core values and goals is related to specific kinds of experiential learning. For example, how would the creation of responsible global citizens be affected by drastic cuts in international programs and service-learning? What other kinds of pedagogies could provide suitable replacements for immersion in a different cultural environment?

ORIENTING PROGRAM REVIEW TOWARD APPLIED LEARNING (AND VICE VERSA)

A comprehensive strategy to embed applied learning in the program review process requires connecting applied learning to performance measures adapted to experiential education. Reconciling applied learning practices with the performance criteria employed in a program review requires a delicate touch. On one hand, an evaluator should not simply re-engineer the criteria to manipulate them in favor of preferred pedagogies. This kind of “working the system” contaminates the process by concealing sub-par performance. On the other hand, the appraisal criteria should have enough breadth and flexibility to encompass the special character of experiential education.

For example, a simplistic calculation of cost per student might quickly identify study abroad as profligate, thereby generating recommendations to curtail or eliminate international components of academic majors. A more nuanced set of criteria, however, would enable per capita costs to be correlated with international student enrollment trends, donations with international program earmarks, and other offsets to costs that a unidimensional per capita measure would miss.

QUALITY CRITERIA

Academic quality often proves elusive, especially if the nature of quality is defined in relation to standards that serve only a limited constituency or rest on fuzzy rationales, such as the momentary impressions of students recorded on course evaluations. Some measures of quality seem directly opposed to philosophical underpinnings of applied learning. For example, correlating student success with earned income fails to acknowledge the deliberate choice of students who enter areas of public service at least partially because they object to the acquisitive values embodied by the concentration of wealth. More systemically, judging achievement mainly by measuring individual accomplishment veers away from the lesson many students glean from service-learning:

Table 1. Hypothetical Alignment of Core Institutional Goals and Values with Applied Learning Practices

Goal or Value	Corresponding Applied Learning Practices
Intensive research activity (culture of research at all levels)	Undergraduate research
Global citizenship, internationalization	Study abroad
Career preparation, job readiness	Internships, practica
Campus-community ties, community outreach	Service-learning

individual prosperity takes a back seat to the public good (Schwartzman & Phelps, 2002). For these students, success consists more of what good they can do than of what goods they can acquire.

Aside from challenging the premises of quality criteria and revising evaluative standards accordingly, those concerned with applied learning can explore additional dimensions of standard quality measures. One area ripe for revision and expansion is the scope of what qualifies as scholarly or creative accomplishments that determine an academic program's status. Practitioners of applied learning could stress the importance of service components, such as increasing the capacity of community agencies to serve clients (an easily quantifiable contribution of service-learning). Another realm that deserves attention is undergraduate research, which could contribute to performance both as research and as teaching due to the extensive mentorship it requires. Finally, various forms of applied learning could qualify as the means for a program to achieve its aspirations. For example, the route to achieving and maintaining a high national ranking in a field might require embracing experiential learning. Such a condition clearly holds in professional schools that educate practitioners in fields such as nursing, physical therapy, and other clinical practices.

FUNCTION AND DEMAND CRITERIA

The function and demand dimension deals with enrollment and placement trends. This area could recognize important contributions of applied learning, but only if criteria are approached creatively. Typically, function and demand refer to student enrollment. For example, an academic program might earn high marks for inclusiveness if it serves a high proportion of students from marginalized racial, ethnic, or socioeconomic populations. The challenge associated with this classification of criteria is that it may be subject to little control or influence by the academic program itself. If an institution as a whole has a highly homogeneous student population, then the program is constrained by those demographics.

A broader measure of inclusiveness and diversity could include not simply the students themselves, but the diversity of the populations with whom they interact via applied learning. Although study abroad will not change the demographic identity of the student traveler, it definitely adds to the diversity of the populations the institution reaches. One indicator of diversity, therefore, could be the number and range of nations serving as destinations for study abroad—a measure a program certainly can influence even if its own students remain demographically uniform. Similarly, service-learning projects could document the demographics of the constituencies served in community partnerships, thereby demonstrating cultural breadth not limited by the student body.

It is tempting to approach demand only from conventional standpoints: primarily student enrollment in courses and proportions of students in field-related employment after graduation. Other measures of demand specific to applied learning might demonstrate the need for accentuating such pedagogy. Demand also can apply to third-party clientele. Demand for service-learning is perhaps more accurately measured in the numbers of incoming requests by community partners for assistance. If a Service-Learning Office or its campus equivalent tracks and records large numbers of community partners whose requests for student volunteers go unfilled, then this point demonstrates high demand that could warrant expansion of service-learning projects. Even if demand measured by student enrollment appears low, external demand could fuel arguments for expanding instead of contracting service-learning experiences. An equivalent argument could justify expansion of study abroad. Although few students may choose international educational experiences, high external demand (large numbers of requests for students) from foreign universities furnishes some evidence for reconsidering curtailment of such programs.

EFFICIENCY CRITERIA

Efficiency might prove the most troublesome for many types of applied learning. Put simply, much applied learning is resource-intensive, and economies of scale often work against pedagogies that require intense preparation and personal mentoring (Schwartzman, 2001). Undergraduate research seems especially vulnerable to efficiency objections, as it frequently involves individual supervision. Even individualized undergraduate research, however, opens the possibility for a reevaluation of efficiency. If undergraduate researchers receive some sort of wages for their work (such as undergraduate research assistantships), one could argue that such personnel usage is highly efficient. Undergraduates would require far lower pay scales than graduate students or faculty, so their research—although apparently inefficient on an absolute scale of students per faculty supervisor—actually represents a far lower labor cost than equivalent work by more advanced staff would incur. This relative efficiency of undergraduate researchers configures them as a cost-saving device rather than as an expensive instructional extravagance.

Other innovative possibilities arise for leveraging applied learning as a way to reduce costs. Alternative paths to teacher certification, a pedagogical issue addressed in the article by McCarty and Dietz in this issue of *JALHE*, offers promising potential for reducing the number of traditional course sections and reducing the time span between teacher recruitment and deployment in a classroom (Coplin, 2005). Applied

learning might fare better than expected according to efficiency criteria if experiential pedagogies are examined more as multidimensional investments than as sheer costs. More concretely, evaluators should explore not only the resources expended to engage in applied learning, but also weigh the costs saved or funds generated by selecting these approaches compared to alternatives.

MICRO-LEVEL: AN ASSESSMENT TOOL TO EVALUATE APPLIED LEARNING

This section discusses the development and testing of an assessment instrument specifically devised to measure educational quality along dimensions salient to experiential education. This assessment tool can be employed across academic disciplines to evaluate the quality of applied learning throughout a campus. The rationale for developing such an instrument arises from the need for generating performance data that emerge from characteristics native to applied learning. If an assessment tool can accurately and reliably identify the quality of applied learning on its own terms, then it allows for comparative evaluations of pedagogical practices throughout the institution. Furthermore, using performance indicators indigenous to applied learning avoids superimposing inappropriate criteria that could fail to capture the most vital pedagogical features of experiential education.

COMMON QUALITY INDICATORS

If one's goal is to build a campus-wide applied learning assessment tool, identifying quality indicators that work across a variety of experiential settings is required. Scholarship focused on developing service-learning as a core pedagogy in higher education has already made great strides in this matter. Eyler and Giles (1999) identify Reflection, Placement Quality, Application, and Diversity as important elements of high-quality service-learning experiences. These elements also apply to determining the quality of internships and practica (Hook & Fern, 1983; Landerholm, Gehrie, & Hao, 2004; Sherman, 1999), undergraduate research (Kinkad, 2003; Mabrouk, 2003), and study away experiences (Cash, 1993; Gillespie, Braskamp, & Braskamp, 1999; Hill & Woodland, 2002).

Reflection in service-learning is the uniting of student experience to academic content through discussion and writing (Eyler, Giles, & Schmiede, 1996). Reflection is relevant for other applied learning forms that have both academic content and student experience as structures. For example, imagine a study away experience in which the intention is that students travel in Rome to study ancient civilization. One version of

this applied experience might involve students and faculty seeing sights together in mornings, with afternoons devoted to contextualizing their experiences through reflection. Simply sightseeing without the appropriate reflection processes (journaling, group discussions, term papers) would be a far weaker learning experience.

Placement Quality in service-learning refers to variety of work, significance of work, and independence in the setting (Eyler & Giles, 1999). Variation along these dimensions exists in other forms of applied learning and could be used to define quality. Consider an internship experience in which a student works on a variety of projects, some of which involve ownership and independence, ultimately yielding products that are presented to administration for consideration. Contrast that internship to the experience a student has when he or she spends a lot of time at a workplace, but has little ownership in projects and no outcome to show for the experience. The former internship is of high placement quality; the latter is weak on this dimension.

Application in service-learning is the degree to which the academic, or disciplinary, content has to be used during the experience (Eyler & Giles, 1999). Yet its usefulness in assessing the quality of a student experience is not limited to service-learning experiences. Any quality applied learning experience relies on application of material in context. For example, an undergraduate research student who participates in the data collection process has to employ content learned in a methods class. This experience has a stronger application element than a research student who merely enters data (rather than analyzing it) or washes test tubes in a laboratory.

Finally, Diversity in service-learning is exposure to a variety of ideas and/or people (Eyler & Giles, 1999). Like the other elements of quality, diversity is arguably significant in a variety of forms of applied learning. Clearly, exposure to cultural diversity through international travel is important to high quality study away experiences. When diversity is broadly construed, though, rather than race- or gender-driven, it is also useful for assessing the quality of internships, practica, and research experiences. For example, an interdisciplinary research team exposes students to diverse perspectives and possibly diverse methods.

To develop an instrument useful as a common assessment tool for a variety of forms of applied learning, the present research builds on the flexibility of these quality indicators identified by Eyler and Giles (1999). Intuitively, and in extant literature, the constructs used to evaluate quality in service-learning apply to other forms of applied learning. To examine this intuition and indication in the literature, this initial study reports the development of a scale around these constructs. The study begins with item development based on a qualitative interview. The interview informs the development of these constructs into a Likert-type

instrument that is quick and efficient to use as part of an institutional assessment program. The research on the instrument itself relies on quantitative methodology to speak to the reliability and validity of the Applied Learning Assessment Tool (ALAT).

VALIDATION STUDY PHASE I

QUALITATIVE INTERVIEW FOR ITEM DEVELOPMENT

Although the constructs discussed above seem applicable to assessment of forms of applied learning beyond service-learning, it is important to put this sort of intuitive analysis to empirical test. Having identified them as potentially useful, the item development phase of this work involved conducting interviews with faculty supervisors of applied learning experiences and holding follow-up focus groups to review drafts of the items resulting from the interviews. This process is summarized below.

Faculty interviews involved sets of questions constructed around Placement Quality, Application, Reflection, and Diversity. Faculty from a mid-size public institution in the Midwest who were identified by their department chairpersons as involved in the applied learning pedagogies were invited to complete the interview process. Of 64 faculty contacted, 53 participated (83% response rate). Interview questions focused on eliciting exemplars of how faculty interpreted each of these constructs in their particular disciplines' applied learning experiences. Interviews were conducted face-to-face and faculty responses were recorded by the interviewer. Each interview took approximately 60 minutes. The interview protocol is available upon request from the fourth author.

Once the interviews were completed, the two interviewers reviewed responses to identify content that might be useful in developing item content. For a response to be considered useful as potential item content, 25% or more of the faculty respondents had to have mentioned it. For example, 54% of faculty agreed that variety, in some form, was essential to their own evaluations of the Placement Quality their students experienced. Thus, at least one or more items assessing "variety of work in the Applied Learning setting" was developed.

Once a set of items was developed, faculty who had participated in the interview process were invited to one or more focus groups. Four focus groups were held, one for each of the four forms of applied learning (service-learning, study away, undergraduate research, and internships/practica). At the focus groups, faculty attendees provided feedback on item content, and the items were further refined. The resultant items comprised the first version of the Applied Learning Assessment Tool. This study documents the evidence supporting the psychometric proper-

ties of the tool. The goal was to develop an assessment tool that could be used for all four forms of applied learning with internal consistency reliability in excess of .80 and a factor structure correspondent to the four common constructs (Placement Quality, Application, Reflection, and Diversity), which would support the construct validity of the scale.

VALIDATION STUDY PHASE I METHOD

Participants. Academic chairs were contacted to determine which courses in their department enrolled students in a curricular applied learning experience during the Fall semester. Students enrolled in identified courses were surveyed. Of the 1,065 questionnaires distributed, 622 were returned (58% response rate).

Instrument. The initial version of the Applied Learning Assessment Tool (ALAT 1.0) contained 17 items, constructed on 5-point Likert scales ranging from 5 (*strongly agree*) to 1 (*strongly disagree*). These items, and their intended subscales, are shown in Table 2. Four parallel versions of the assessment measured the four types of applied learning experiences: undergraduate research/project, internship/practicum, service-learning, and study away. The four parallel versions allowed for slight wording variations to make items clearer to the student. For example, item 4 reads: "This internship/practicum experience increased my exposure to different types of people." The only change from each version was the experience name, such that each version read, "this independent research/project," "this study away," or "this service-learning experience."

VALIDATION STUDY PHASE I RESULTS

Reliability. The 17 items across the four types of applied learning experiences showed adequate internal consistency reliability as measured by Cronbach's alpha ($\alpha = .88$). Because this survey was developed to support institutional accreditation processes, it was intended to be used each semester. Because of the anticipated frequency of use, one goal was to create a more concise version of the tool.

To reduce length, item-to-total correlations were examined to determine which items were good candidates for omission from the scale, while maintaining an internal consistency reliability of .80 or greater for the global scale. Item-to-total correlations show how individual items correlate to the global scale. A low correlation indicates the item may measure a different construct than other items. This analysis also shows how the global alpha will change when a given item is removed. The aim was to remove items with lower item-to-total correlations, yet retain

Table 2. ALAT 1.0 Items

Item	Subscale
1. Could have had same experience without coursework	Application
2. Involved in multiple steps	Placement Quality
3. Did not have the opportunity to function independently	Placement Quality
4. Increased my exposure to different types of people	Diversity
5. No writing was required of me	Reflection
6. Supervising professor provided no feedback	Reflection
7. My contribution was significant to the outcome	Placement Quality
8. I problem solved	Reflection
9. Participated in action planning	Reflection
10. Did not see a variety of viewpoints expressed	Diversity
11. Not exposed to a diverse group of people	Diversity
12. Supervising professor and I discussed the significance	Reflection
13. Relates to other classes	Application
14. Syllabus/assignments did not connect	Application
15. Relates to my coursework	Application
16. Increased my exposure to viewpoints different than my own	Diversity
17. Spent the majority of the time doing low-level, repetitive tasks	Placement Quality

a relatively balanced subscale structure such that each subscale had at least three items. Based on item-to-total correlation analysis, the following items were removed: (1) Item 7, from Placement Quality subscale, (2) Item 13, from Application subscale, (3) Items 5 and 6 from Reflection subscale, and (4) Item 11, from Diversity subscale. A total of 12 items remained, achieving an alpha of .864. These remaining 12 items will be referred to as ALAT 2.0 hereafter.

Construct Validity. As discussed previously, ALAT 2.0 comprised four distinct subscales: Placement Quality, Application, Reflection, and Diversity. To determine if this hypothesized factor structure could be empirically verified, a confirmatory factor analysis was conducted. Because the four subscales were anticipated to correlate with each other, an oblique rotation (Promax) using SPSS was selected.

The result of the 4-factor structure accounted for 64% of the variability in the data. The four factors, after rotation, were roughly consistent with the hypothesized factor structure. Factor loadings are shown in Table 3. Two criteria were used in evaluating the factor structure: (1) items should load on the same factor as items in the same subscale and (2) loading on this factor should be at least .40 (Gorsuch's [1983] minimum). Each subscale's performance in the analysis is reviewed in Table 3.

Application subscale items performed the best, with all items loading at least .40 or higher on the same factor. Reflection items fared reasonably well, with all items loading at least .40 on the same factor. One item also loaded on the Application factor. Diversity subscale items all loaded at least .40 on the same factor as well with one of the three items also loading on the Application factor. The Placement Quality subscale was weakest. Of the three items hypothesized to be Placement Quality items, only one loaded on the Placement Quality factor. The remaining two items loaded on the Application factor.

To better understand the poor performance of the Placement Quality subscale, item content was examined. The two Placement Quality items

Table 3. Factor Pattern for Validation Study Phase I, Fall Data Set

Subscale	Factor			
	Placement Quality	Application	Reflection	Diversity
<i>Placement Quality</i>				
2. Involved in multiple steps.	.99			
3. Did not have opportunity to function independently. (R)	.34	<i>.40</i>		
17. Spent majority of time doing low-level tasks. (R)	.37	<i>.47</i>		
<i>Application</i>				
1. Could have had same experience without coursework at this university. (R)		.59		
14. Assignment did not connect. (R)		.75		
15. My research/project relates to major coursework.		.68		
<i>Reflection</i>				
8. I problem solved.			.81	
9. I participated in action planning.			.84	
12. My supervising professor and I discussed significance of research/project.		<i>.54</i>	.47	
<i>Diversity</i>				
4. This increased my exposure to different types of people.				.49
10. I did not see a variety of viewpoints. (R)		<i>.53</i>		.42
16. This increased my exposure to viewpoints different than my own.				1.00

NOTE: (R) indicates a reverse-score item. The items in this table appear in an abbreviated form. Boldfaced numbers indicate factor loadings on which the item was predicted to load; italicized numbers indicate a factor loading above .40 that was unanticipated for that item.

that loaded on the Application factor were intended to measure independence (Item 3) and variety (Item 17) within the applied learning placement. This may indicate that Placement Quality as a subscale is more conceptually complex than the other three subscales. If the Placement Quality subscale measured more than one construct, it could account for the pattern of item loadings, indicating multiple subscales are needed to address Placement Quality. Alternately, the weak pattern loading could simply be a result of poorly worded items. A second wave of data collection provided an opportunity to revise these items.

VALIDATION STUDY PHASE I DISCUSSION

Although the instrument's reliability went down slightly from ALAT 1.0 to 2.0, the scale was shortened while retaining an acceptable global alpha. Removing items with low item-to-total correlations allowed creation of a more succinct and streamlined assessment that presumably assesses all four types of applied learning experiences while keeping intact the reliability of the measurement.

The factor structure showed Placement Quality items did not load onto a distinct factor, instead partially loading on the Application factor. This may have indicated the Placement Quality items were not statistically measuring the same construct, and suggests that multiple subscales might be required to adequately assess Placement Quality. However, before adding additional subscales, which would increase the length of the scale, the alternate possibility that the existing Placement Quality items were simply not well written was tested.

Loadings on the factor structure indicated that items in the Application subscale statistically measured the same construct, as they all loaded on the hypothesized factor. Items in the Reflection and Diversity subscales loaded into almost distinct factors, with only one item in each subscale loading onto the Application factor. The hypothesized Application factor clearly contained more than the expected three application items, indicating it may be difficult to disentangle from the other subscale items. Given the overall construct is applied learning, this is not remarkable.

Finally, the present study provided no evidence of convergent validity. If evidence of convergent validity were demonstrated in the following study, the ALAT would be a stronger assessment tool for institutions to use when evaluating student applied learning experiences.

VALIDATION STUDY PHASE II

VALIDATION STUDY PHASE II METHOD

Participants. The same method of distribution used in the Fall data set was used for Spring. A total of 1,190 questionnaires were distributed, with 709 returned (59.6% response rate).

Instrument. The ALAT 2.0's Placement Quality items were edited for clarity in an attempt to improve subscale performance. The resulting scale will be referred to as ALAT 3.0 hereafter, and can be seen in Table 4. The ALAT 3.0 was used in the present study. As before, all items employed a 5-point Likert scale. The 12 items comprised four hypothesized scales: Placement Quality, Application, Reflection, and Diversity.

Study 1 offered no evidence of convergent validity. Four items additional to the 12-item scale were included in ALAT 3.0, aimed to address this weakness. The four validity items were: (1) I am satisfied with my experience, (2) I would recommend this experience to a friend, (3) I would repeat this experience, given the opportunity, and (4) This experience seemed of little value. The first three items were predicted to correlate significantly and positively with overall quality scores as

Table 4. ALAT 3.0 Items

Item	Subscale
1. Could have had same experience without my coursework	Application
2. Involved in multiple tasks	Placement Quality
3. Had to function independently	Placement Quality
4. Increased my exposure to different types of people	Diversity
5. I problem solved	Reflection
6. Participated in action planning	Reflection
7. Did not see a variety of viewpoints expressed	Diversity
8. My professor led discussions about the significance	Reflection
9. Syllabus/assignments did not connect this experience to my other classes	Application
10. Relates to my major coursework	Application
11. Increased my exposure to viewpoints different than my own	Diversity
12. Spent the majority of my time performing the same type of task repeatedly	Placement Quality

measured by ALAT 3.0, while the fourth item was expected to correlate significantly and negatively with the same.

VALIDATION STUDY PHASE II RESULTS

Reliability. ALAT 3.0 showed moderate internal consistency reliability, measured by Cronbach's Alpha ($\alpha = .78$). This was a much lower reliability than Study 1; however, the ALAT 3.0 contained fewer items, making higher reliability more difficult to demonstrate. Because the internal consistency reliability was lower than expected, the item-to-total correlations were re-examined to determine if any particular items were driving the alpha down. In particular, the edited Placement Quality items were scrutinized. As indicated above, Placement Quality items were anticipated to potentially perform poorly simply because they measure multiple constructs. These edited items may have contributed to the initially weaker alpha observed.

Items showing markedly lower item-to-total correlations were removed. Items 3 and 12—both from the Placement Quality subscale—were removed because, as suspected, the alpha for the global scale increased with their removal. These were the revised versions of the poorly performing Placement Quality items from Study 1. The remaining 10 items (ALAT 4.0) were then re-entered into internal consistency analysis, improving alpha to .81.

Construct Validity. To test the subscale structure, the remaining items were submitted to a confirmatory factor analysis. Because there was only one Placement Quality item (item 2) remaining in ALAT 4.0, only three complete subscales remained: Application, Reflection, and Diversity. Therefore, only three factors were expected. The remaining Placement Quality item essentially measured variety of activity, which might also function as part of the diversity of the experience. Consequently, this item was predicted to load on the diversity factor, as that assessed variety within the student experience.

The same oblique rotation method (Promax) was used, with a 3-factor solution specified. The factor analysis accounted for 58% of the variability in the data. Factor loadings are shown in Table 5. As before, the same two criteria from Study 1 were used to evaluate factor structure.

All three of the Application subscale items loaded on the hypothesized factor over .40 with one item also loading on the Diversity factor. The Reflection subscale had two items load onto the predicted Reflection factor, with one item loading above the .40 threshold on the Diversity factor rather than the Reflection factor. The Diversity subscale had four hypothesized items, including the former Placement Quality item (Item 2). As expected, the former Placement Quality item loaded on the

Table 5. Factor Pattern for Validation Study Phase II, Spring Data Set

Subscale	Factor		
	Application	Reflection	Diversity
<i>Application</i>			
1. Could have had same experience without coursework at this university. (R)	.65		
9. Syllabus/assignment did not connect. (R)	.79		
10. Relates to major coursework.	.68		<i>.80</i>
<i>Reflection</i>			
5. I solved problems.		.77	
6. I participated in action planning.		.76	
8. My professor led discussions about significance.		.17	<i>.64</i>
<i>Diversity</i>			
2. Involved in multiple tasks.			.71
4. This increased my exposure to different types of people.			.71
7. I did not see a variety of viewpoints. (R)		<i>.76</i>	.22
11. This increased my exposure to viewpoints different than my own.			.79
NOTE: (R) indicates a reverse-score item. The items in this table appear in an abbreviated form. Boldfaced numbers indicate factor loadings on which the item was predicted to load; italicized numbers indicate a factor loading above .40 that was unanticipated for that item.			

Table 6. Correlations between 10-item ALAT and Convergent Validity Items for Validation Study Phase II, Spring Data Set

	Validity Item			
	1. Satisfied	2. Recommend	3. Repeat	4. Little Value
10-item scale (ALAT 4.0)	.66	.28	.34	-.64
Validity Item 1 (Satisfied)		.88	.62	-.72
Validity Item 2 (Recommend)			.20	-.32
Validity Item 3 (Repeat)				-.36

Diversity factor. Two of the remaining three Diversity items also loaded on that factor, with the final item loading well below the .40 threshold. Rather, it loaded on the Application factor, as it did in Study 1.

Convergent Validity. Beyond refining the subscale structure, a second goal of Study 2 was to establish convergent validity of the ALAT. The four validity items were entered into a correlation with the mean score for the 10-item scale. Correlation coefficients are shown in Table 6. All four items, as expected, correlated significantly at the .01 significance level and in the predicted directions.

VALIDATION STUDY PHASE II DISCUSSION

Moving from ALAT 3.0 to 4.0, adequate internal consistency reliability was achieved. When the two poorly performing Placement

Quality items were removed, not only did the scale's alpha improve, but the result was an extremely brief assessment tool useful across applied learning types, a goal from Phase I. The removed items performed poorly in both Phases I and II. When eliminated, internal consistency improved, indicating these items were measuring constructs somewhat different than "applied learning quality."

With only one Placement Quality item remaining, a complete Placement Quality subscale no longer existed. Indeed this possibility was anticipated, as Placement Quality is a complex construct to measure in a single subscale. The remaining Placement Quality item was reassigned to the Diversity subscale, and a 3-factor structure was expected. As anticipated, the Placement Quality item loaded onto the diversity subscale. This was reasonable as the former Placement Quality item measured variety within the placement site, and was similar to the diversity construct.

Application fared well in Phase II of the Validation Study, with all items loading above threshold on the hypothesized factor. One item dually loaded on the Diversity factor. Reflection performed adequately, with two items loading as predicted, and one item loading unexpectedly on the Diversity factor. The Diversity subscale had three items load on the Diversity factor, and one item loaded on the Application factor. This was the same item that loaded onto the Application factor in Phase I.

Both Application and Reflection showed an item being loaded on the Diversity factor, indicating Diversity, like Placement Quality, might encompass more than originally considered. Further, Diversity had one item that loaded on the Application factor. Although adequate distinction was achieved between these two factors, future research should focus on rewording such items to more narrowly define each subscale construct, with less overlap between them.

Another goal for Phase II of the Validation Study was to provide evidence of convergent validity for the ALAT. This was met with strong results. As expected, the four items used to establish this validity significantly correlated with the global scale score in predicted directions. Future research might strengthen the scale further by seeking evidence of divergent validity, and even more important, criterion-related validity.

GENERAL DISCUSSION OF THE INSTRUMENT

The primary objective was to create a succinct assessment tool that could be used for institutional purposes to measure applied learning experiences. Because each applied learning experience shares common elements of quality (Application, Reflection, and Diversity), using a single assessment tool improves comparability. The current version's 10-item length makes it amenable for use each semester.

The reliability and validity of the ALAT 4.0 are other strengths that should appeal to those interested in using the instrument for institutional assessment purposes, or for individual faculty/staff development feedback. The Fall data set revealed strong internal consistency, but included items that failed to measure the hypothesized subscales according to the factor analysis. Moving to Spring semester, the number of items was pared, while maintaining strong internal consistency reliability. The ALAT 4.0 had only 10 items, meeting the goal of a concise tool, while preserving a strong alpha. Factor analyses from both Phases I and II indicated reasonable subscale validation. Items in each subscale loaded onto its hypothesized factor a majority of the time, indicating that the hypothesized subscales are useful measures for a variety of forms of applied learning. Finally, convergent validity evidence was very strong from Phase II, with all items correlating in predicted directions and magnitudes with the overall scale score.

LIMITATIONS AND IMPLICATIONS

Although the subscales showed evidence of construct validation, factor analyses also revealed some items loading into unpredictable patterns. For instance, in Phase II, Reflection and Diversity each had one item load into a different factor than anticipated. Given that subscales would reasonably be expected to correlate, it is likely there is some overlap in items predicted to load onto separate factors. Placement Quality loaded in unexpected ways with even more frequency than items predicted to load onto the other three subscales. This consistent "double-loading" indicated Placement Quality, as it was defined in these studies, was likely too broad to be assessed in a single subscale. Placement Quality comprises several ideas: variety, independence, and diversity within the placement. This breadth of content could explain why Placement Quality items kept loading on Application and Diversity factors. Future research should focus on establishing these additional subscales.

Although the convergent validity evidence was strong in Phase II, no evidence of divergent or criterion-related validity was provided in these studies. The focus on content (item development phases), construct, and convergent validity, while useful, is not a complete picture of the assessment tool. Future studies can fill this gap by identifying appropriate measures for establishing divergence and criterion-related validity. For example, within a single course that contains applied learning experiences, a professor might be able to rank order his or her perceived quality of students' experiences. These rankings could then be correlated with quality as assessed by the ALAT 4.0 to establish a concurrent criterion-related validity coefficient.

The ALAT 4.0 is capable of assessing disparate programs using a single tool. This makes evaluations and comparisons of programs that fall within the applied learning rubric more feasible. Beyond evaluation, ALAT 4.0 can highlight common strengths and weaknesses across programs, allowing administrators to plan professional development activities to maximize cross-program impact.

Kolb (1981) has articulated a well-known theory of experiential learning, which describes how students process experience and learn from it. Kolb describes students' movement through a learning cycle of concrete experience, reflective observation, abstract thinking, and active experimentation. This theory is useful for understanding a student's development, whether the learning activity is curricular or co-curricular. Kolb argues that effective learners need competencies in all four components of the learning cycle. Students must involve themselves fully and without bias in learning experiences, observe and reflect on those experiences from diverse perspectives, formulate concepts through reflection that integrate observations into personal theories, and apply those theories in decision-making and problem-solving. It is this complex integration that creates a quality learning experience.

Although the ALAT 4.0 was developed to measure curricular applied learning, its three subscales of Diversity, Reflection, and Application correspond nicely to the last three stages in Kolb's (1981) learning cycle. Perhaps part of the Placement Quality subscale's failure, in addition to explanations offered above, is that it failed to correspond well to Kolb's notion of concrete experience. The correspondence between Kolb and the remaining three subscales suggests that, though currently only tested in curricular applied learning settings, the ALAT 4.0 should be useful to student development professionals in assessing co-curricular applied learning experiences. Certainly a strong co-curriculum builds in opportunities to expose students to diverse populations and perspectives, to reflect upon the experience and form an understanding of the system in which the experience is embedded, and then apply that new understanding to future activities. The ALAT 4.0 subscales provide both a guide for student development professionals creating applied learning experiences and an assessment tool to document the quality of the activity to institutional and community stakeholders.

SYNTHESIS: WHAT APPLIED LEARNING AND ACADEMIC PROGRAM REVIEW MEAN FOR EACH OTHER

Development of the ALAT or equivalent home-grown instruments can facilitate not only comparative performance appraisals of applied learning, but can establish local performance benchmarks when external

yardsticks are unavailable. Gathered over time, consistent and wide-spread data garnered with such tools can illuminate longitudinal trends that can provide valuable indicators of relative educational quality over time.

Incorporating rigorous program review into an institution's culture furnishes much clearer evidence for claims about the nature and quality of academic programs. More important, program review can generate ongoing, detailed feedback about how each program is progressing toward its aspirational goals. The more proactive, ongoing use of program review is diagnostic: noticing patterns of performance that can justify the choices the denizens of academia make.

A multidimensional program review can reveal aspects of academic identity in ways that go beyond simply counting bodies in seats or weighing reams of publications. Yes, those things do matter, but they are parts of larger wholes that constitute a holistic education with identifiable indicators of quality. More important, when legislators, parents, prospective students, faculty at other campuses, or other interested parties ask about academic programs, an institution can cite specific, documented rationales for its decisions about resource allocation. Program review, used proactively, can provide program preview: clearer insight regarding what a program can and should be. An especially important and uplifting aspect of program review deals with the implications of identifying the highest performing academic programs. Program review involves collecting data, history, and cogent rationales that show where an institution excels and why. Academic leaders can make a much better case for promoting the institution, soliciting funds, and recruiting because their claims of excellence have much more tangible evidence. This essay has attempted to demonstrate the roles that applied learning could play in making those claims about academic performance—and academic possibilities.