Blinded Abstract Body

Lip Service or Actionable Insights? Linking Student Experiences to Assessment, Accountability, and Data-Driven Decision Making in Higher Education

Purpose

Over the last three decades, the general public and policymakers have become increasingly focused upon the issues of accountability and data-driven decision-making within the postsecondary context (Altbach et al., 2005; Paulsen & Smart, 2001). More recently, subtle changes in the landscape of higher education have brought a new focus to accountability and with it "increasing pressure from various constituencies to demonstrate its effectiveness in measurable terms" (p. xi). Traditionally, the focus of assessment in higher education has been on student outcomes. More recently, however, assessment activities have broadened to include resource allocation, accreditation standards, comparisons with peer institutions, and satisfaction of students, faculty, and administrators (Ory, 1992; Shuh & Upcraft, 2001).

The arguments supporting or challenging this movement toward accountability (and its cousins "assessment" and "data-driven decision-making") often reduce to polarized and politicized ideological beliefs about the purpose of higher education and the mechanisms through which the academy can maintain (or develop, depending on your point of view) educational excellence and operational efficiency. Nonetheless, underlying the argument on all sides is the assumption that educational quality is likely to be improved when decision-makers (at all levels) develop policies and implement practices informed by relevant assessment data. Assessment is considered the vehicle to determine which of those policies and practices lead to improved student experiences, greater learning, and/or higher satisfaction. However, this assumption regarding the value of accountability, assessment, and data-driven decision-making remains largely untested.

Therefore, we use empirical data from senior administrators at 57 diverse institutions across five states to examine the extent to which assessment and data driven decision making shape the experiences of nearly 9,000 first-year students. Specifically, this paper addresses two main questions:

- 1. Which institutions (and which divisions therein) are employing which types of assessment and data-driven decision-making regarding students' first year of college?
- 2. To what extent does institutional adoption of assessment and data-driven decisionmaking correspond to levels of first-year student engagement and learning gains?

Background and Context

The literature on the impacts of assessment and evaluation on organizations tends to take one of two distinct normative positions. The "accountability myopia" thesis (Ebrahim, 2005, p. 56) warns of the unintended consequences of assessment practices in organizations while the "reflective practice" thesis (Schon, 1983, p. 1) argues that assessment activities are a necessary ingredient for organizational learning.

Some of the assessment literature takes the normative position of "accountability myopia" thesis (Ebrahim, 2005, p. 56). There are two reasons that accountability can be viewed as myopic. First, assessment may often represent a purely symbolic function in higher education organizations (Feldman & March, 1981). Thus, "evaluations can be undertaken for the symbolic purpose of legitimating existing activities rather than for identifying problematic areas for

improvement" (Ebrahim, 2005, p. 67). The second reason that accountability can be regarded as myopic is that outcome measures important to powerful stakeholders are more likely be assessed than outcome measures important to less powerful groups. For instance, the economic benefits of higher education are widely assessed and subsequently reported by the popular press and business groups whereas the psychosocial benefits of college such as prejudice reduction or civic engagement are less frequently assessed and may only be monitored by academic researchers and community activists. Thus, "accountability is also about power, in that asymmetries in resources become important in influencing who is able to hold whom accountable" (Ebrahim, 2005, p. 60). Shore and Wright (2000) refer to this phenomenon in higher education as "coercive accountability" (p. 57).

In contrast, the "reflective practice" literature reveals a different normative position (Schon, 1983, p. 1). While reflective practice often refers to the behavior of individuals, this term can also be applied to institutions as a whole. The reflective practice literature suggests that organizational learning can result from assessment practice, but only when knowledge is used to change behavior (Argyris & Schon, 1996; Astin, 1993; Ebrahim, 2005; Senge, 1990).

Even those who draw hope from the more optimistic position of reflective practice concede that institutions of higher education are slow-moving entities in which the collection of assessment data collection far outpaces institutions' use of results to inform policy or practice. As Banta and Blaich (2011) acknowledge,

"We scoured current literature, consulted experienced colleagues, and reviewed our own experiences, but we could identify only a handful of examples of the use of assessment findings in stimulating improvements. In fact, among 146 profiles of good practice submitted by colleagues at campuses from across the country for possible inclusion in a new book, Trudy Banta, Elizabeth Jones, and Karen Black found that only 6 percent of the profiles contained evidence that student learning had improved, no matter what measure had been used. Likewise, in their evaluation of the Wabash National Study, Charles Blaich and Kathleen Wise noted strong campus engagement with the process of assessment but few instances of actual change in response to the information generated by the study." (p. 22)

In describing several barriers to effective assessment, Banta and Blaich (2011) implicitly support the idea that data-driven decision making should could be "closing the loop" (p. 23) between data collection and institutional change, thereby enabling institutions to create or maintain environments conducive to student engagement and learning.

Method

The current study uses empirical data from 114 senior administrators and nearly 9,000 students at 57 diverse institutions across five states to examine the extent to which institutional assessment and data driven decision making shape the experiences of first-year college students.

Data Sources. This paper draws its data from 57 bachelor's degree granting institutions across five states (California, Florida, Iowa, Pennsylvania, and Texas). In Spring 2012, project staff sent printed copies of the *Survey of Academic Policies, Programs, and Practices* to the Chief Academic Officer (typically the Provost); the *Survey of Student Affairs Policies, Programs, and Practices* was likewise distributed to the Chief Student Affairs Officer (typically the V.P. for Student Affairs or Dean of Students). Together, these surveys' items form *a-priori* scales reflecting clusters of institutional policies (e.g., student affairs emphasis on diversity, first-

year seminars, and recent assessment efforts; see Table 1 for details). Analyses reported in this paper make use of 5 such policy-cluster scales (alpha between 0.79 and 0.90) and one meta-scale (alpha of .74) that incorporates these same five scales plus seven others to describe the extent to which an institution was intentional in its use of administrative policies and structures to shape first-year students' engagement in educationally purposeful activities.

Each scale reflects the extent to which institutions have aligned their policies with the broad conclusions from the currently available research literature on first-year student success. A scale score of one indicates an institution's complete adoption of all measured policies related to a particular policy cluster. It may be most useful to think of the scale scores as approximations, where a score of 0.5 indicates that an institution has done approximately one-half (50%) of what it could be doing to align its policies with the available research.

Student data come from the National Survey of Student Engagement (NSSE) completed by 8,847 first-year students attending the 57 participating institutions. All student-level data are weighted to reflect the gender and full-time/part-time distribution of first-year students at each participating institution. Although self-reported data in general, and NSSE specifically, have received some recent criticism (Bowman, 2011; Campbell & Cabrera, 2011; Porter, 2011), the use of NSSE scales in this paper is appropriate because a) we draw our data from multiple institutions; b) the random intercept multi-level model, as we use it, addresses variations in the conditional mean scale score for whole institutions, not individual students; c) we do not rely on any single item, or even any single scale, to draw conclusions.

Variables: Dependent/criterion variables in these analyses are all scales composed of items from the National Survey of Student Engagement (see Table 2). For these analyses, we consider the five traditional NSSE benchmarks of effective educational practice in undergraduate education, two measures of "deep" learning experiences, and three self-reported measures of learning and development during the first year of college. To isolate the effects of institution-level policies, we included several control variables at the student level. Following Raudenbush and Bryk's (2002) guidelines, we centered all level-1 (student) control variables around their grand means. Each model includes variables representing students' gender/sex, race/ethnicity, age, major, first-generation status, transfer status, full/part-time enrollment, involvement with varsity athletics, on-campus residence, and their ACT Composite score. We also entered several uncentered level-2 (institution) control variables. Level-2 control variables reflect institutional characteristics not easily or quickly manipulated by colleges or universities: public/private, highest degree offered, undergraduate enrollment, racial distribution of student body, Pell grant eligibility of the student body, and admissions selectivity as rated by the Barrons guide.

Missing Data Augmentation: Data were augmented using a multiple imputation procedure whereby a new dataset was drawn after each 100 iterations of the imputation model to create 10 distinct dataset. Analyses were run independently on each of the 10 imputed datasets, with results then pooled across the 10 analyses.

Analytic Approach: Analyses proceeded through a series of multi-level random intercept (a.k.a. "fixed effects") models with an unrestricted variance/covariance matrix. Because our independent variable of interest occurs at level-2 (institution), where we have just 57 institutions, we use the critical p-value of 0.10 instead of the more traditional 0.05.

Power, Robustness, and Interpretation: We have taken several precautions to avoid threats related to both type 1 and type 2 errors of interpretation. At no point do we base our interpretation on a single "statistically significant" coefficient or an arbitrarily chosen cut point for reduction in level-2 residual variance. Moreover, we acknowledge that none of the results

would independently remain statistically significant were we to apply individual Bonferreni posthoc adjustments for critical p-values. Rather than rely on any single statistical result, we look for patterns of results that are suggestive of underlying policy effects. Together with the use of multiple-imputation and multi-level modeling, these interpretive cautions specifically protect against type-1 errors (incorrectly claiming a policy effect). Moreover, we protect against type-2 errors (incorrectly dismissing a true policy effect) in two ways. First, power analyses (using the Optimal Design software specific to multi-level models) indicate that our data and analyses can, with 90% power, detect effect sizes of .15 to .22 (the precise minimum detectible effect size varies with the intra-class correlation of each dependent variable). Moreover, we never make a claim of "no policy effect" based on a single statistical result. Rather, our claims regarding a lack of policy effect are based on the absence of a statistically significant policy effect for *all* of the 10 outcome variables and across at least two scales reflecting related policy clusters.

Findings

Descriptive statistics from our study suggest that institutions are starting to embrace recent calls to increase the use of assessment and data-decision making. Mean scale scores indicate that institutions have adopted roughly 49-69% of this study's literature-supported policies related to assessment and data-driven decision making. Moreover, examination of the scales' standard errors, as well as individual items from the surveys (not reported here), reveal meaningful variation across institutions. For example, although two schools in our study have adopted 100% of the student affairs assessment policies we measured, two other schools have adopted none of those policies.

However, counter to researcher expectations, policies related to a) assessment, b) datadriven decision making, and c) institutional intentionality do not appear related to student engagement or self-reported gains during the first year of college. As table 2 shows, although we analyzed a total of 60 models (6 policy scales x 10 student outcomes) and used a liberal critical value for statistical significance (0.10), in *none* of the analyses did *any* the policy scales coefficients indicate a statistically significant relationship with *any* of the project's ten dependent/criterion variables.

Although methodological limitations and scholarly caution make us reluctant to draw conclusions based on any individual model's non-significant findings (for fear of making a type-2 error), the overarching pattern of non-significant findings runs counter to researcher expectations and differ dramatically from results of other policy analyses from the same project. When replicated using other policy scales from the same dataset, models using 21 other policy scales yield 44 instances of statistical significance (results reported elsewhere). Moreover, our current findings of non-significance even defy expectations of simple random chance. With a critical p-value of 0.10, researchers would expect to make a type-1 error (finding statistical significance and incorrectly claiming a policy effect) in approximately 10 percent of analyses. Applied to the current study, we would expect our 60 analyses to produce 6 statistically significant policy coefficient suggest that our findings of non-significance are interpretable as reflecting the overall ineffectiveness of policies related to assessment, data driven decision making, and administrative coordination of the first year experience.

Conclusions and Implications

It is easy to see how the contrasting perspectives of reflective practice and accountability myopia

might drive both the desperate hope and intense fear often expressed by different stakeholders when discussing accountability, assessment, and data-driven decision making in higher education. It is the *hope* of "reflective practice" (Schon, 1983, p. 1) leading to educational improvement that undergirds the push (typically by policy-makers) for institutional accountability and (by some higher education scholars) the development of a "culture of assessment" in higher education. But the *fear* of "accountability myopia" (Ebrahim, 2005, p. 56) unsettles many faculty members and fuels resistance to more intense federal involvement in postsecondary accreditation or the adoption of No-Child-Left-Behind-like mandates within higher education.

Our findings suggest that, whatever their perspective, college and university administrators have started to get the message about the coming "age of accountability" in higher education. Nearly all of the 57 schools in this study regularly collect some form of assessment data; more than half of those same institutions report using assessment data to inform decision making about personnel, courses, programs and/or resource allocation. However, counter to research expectations, the institutional adoption of policies related to the collection of assessment data or the application of data-driven decision making appears to have no relationship with student experiences or outcomes in the first year of college.

These findings, though perhaps discouraging or counterintuitive, are consistent with the nascent body of literature questioning the effectiveness of accountability and assessment policies in higher education. At the state level, Tandberg and Hillman (2013) found that, "while performance funding may have brought forth other outcomes not examined in our studies (e.g., greater accountability and oversight), it has generally not achieved the most basic goal all states believe is central to their performance efforts—improving degree productivity" (p. 7). At the institution level, Banta and Blaich (2011) have noted how difficult it is to find evidence that assessment efforts have led to improved student learning. In both cases, the authors hold some hope that, with sufficient tweaking, state accountability and institutional assessment efforts can lead to improved student outcomes. But results from the present study seem likely to validate more than placate any lingering fears of the "accountability myopia" Ebrahim (2005, p. 56) warns about.

Appendix A - References

Not included in page count.

- Altbach, P. G., Berdahl, R. O., & Gumport, P. J. (Eds.). (2005). *American higher education in the twenty-first century: Social, political, and economic challenges* (2nd ed.). Baltimore: Johns Hopkins University Press.
- Argyris, C. & Schon, D.A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley.
- Astin, A.W. (1993). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education. Phoenix, AZ: Oryx Press.
- Banta, T. W., & Blaich, C. (2010). Closing the Assessment Loop. *Change: The Magazine of Higher Learning*, *43*(1), 22-27. doi: 10.1080/00091383.2011.538642
- Bowman, N. (2011). Examining systematic errors in predictors of college student self-reported gains. *New Directions for Institutional Research*, *150*, 7-19.
- Campbell, C. & Cabrera, A. (2011). How sound is NSSE? Investigating the psychometric properties of NSSE at a public, research-extensive institution. *The Review of Higher Education*, *35*(1), 77-103.
- Ebrahim, A. (2005). Accountability myopia: Losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly*, *34*(1), 56-87.
- Feldman, M.S. & March, J.G. (1981). Information in organizations as signal and symbol. *Administrative Science Quarterly*, 26(2), 171-186.
- Gardner, J.N., Barefoot, B.O., Schwartz, S.W., Siegel, M.J., Swing, R.L., Reason, R.D., Terenzini, P.T., Zlotkowski, E. (2005). Foundational dimensions: First-year focus (fouryear college version). Retrieved from http://www.jngi.org/foe-program/foundationaldimensions/four-year-first-year-focus/
- Gonyea, R. & Miller, A. (2011). Clearing the AIR about the use of self-reported gains in institutional research. *New Directions for Institutional Research*, *150*, 99-111.
- Hox, J. (2010). Multilevel analysis: Techniques and applications. Routledge.
- Kuh, G.D., Kinzie, J., Schuh, J.H., Whitt, E.J., & Associates. (2010). Student success in college: Creating conditions that matter. San Francisco: Jossey-Bass. Ory, J.C. (1992). Metaassessment: Evaluating assessment activities. Research in Higher Education, 33(4), 467-481.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.
- Pascarella, E., & Terenzini, P. (2005). How college affects students: Findings and insights from twenty years of research. Volume 2. A third decade of research. San Francisco: Jossey-Bass.
- Porter, S. (2011). Do college student surveys have any validity? *The Review of Higher Education*, 35(1), 45-76
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear modes: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA. Sage.

- Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Currency Doubleday.
- Schuh, J. H., & Upcraft, M. L. (2001). Assessment Practice in Student Affairs: An Application Manual. The Jossey-Bass Higher and Adult Education Series: ERIC.
- Paulsen, M. B., & Smart, J. C. (2007). *The finance of higher education: Theory, research, policy, and practice*: Algora Publishing.
- Shon, D.A. (1983). *The reflective practitioner: How professionals think in action*. United States: Basic Books.
- Shore, C. & Wright, S. (2000). Coercive accountability: The rise of audit culture in higher education. In M. Strathern (Ed.), *Audit cultures: Anthropological studies in* accountability, ethics, and the academy, pp. 57-89. London: Routledge.
- Tandberg, David and Hillman, Nicholas (2013). State Performance Funding for Higher Education: Silver Bullet or Red Herring? (WISCAPE POLICY BRIEF). Madison, WI: University of Wisconsin-Madison, Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE).

Appendix B - Tables and Figures

Not included in page count.

Table 1: Descriptive Statistics for Policy Scales

	Scale Alpha	Mean	S. D
Assessment of Student Affairs Programs (Student Affairs)	.79	.686	.256
Recent Assessment Efforts: Learning (Academic Affairs)	.81	.521	.432
Recent Assessment Efforts: Persistence (Academic Affairs)	.82	.538	.426
Data-Driven Decision Making (Student Affairs)	.81	.613	.217
Data-Driven Decision Making (Academic Affairs)	.90	.489	.320

Table 2: Results from Multi-Level Models Linking Institutional Policies to Student Experiences and Outcomes

	NSSE Benchmarks ¹				Deep l	Learning ²	ũ	Gains in.	<u>,</u> 3	
_{ອັເລີ} Scale Name / Description	EEE	SCE	AC	SFI	ACL	DPv1	DPv 2	PRC	PSO	GED
^{জ ব} % variance between institutions (intra-class correlation)	7.3%	4.2%	7.0%	2.9%	5.2%	3.5%	3.2%	3.2%	5.5%	2.8%
Overall Policy Adoption Scales		ļ								•
.80 Overall Policies						ļ				\uparrow
.67 Student Affairs Policies				\uparrow						<u> </u>
.67 Academic Affairs Policies										
.74 Institutional Intentionality Composite Scale										
.90 Diversity Composite Scale				\uparrow						<u> </u>
Student Affairs (SA) Scales	ļ									
.89 Programmatic Emphasis on Diversity										^
.85 Info. Dissemination: Frequency of Contact	L		\uparrow	\uparrow	\uparrow			\uparrow		
.83 SA Service Availability		<u> </u>		1						^
.82 Program Emphasis on Socio-Political Issues										
.81 FY Orientation: Specific Student Pop.	\uparrow			\mathbf{T}						<u> </u>
.81 Data-Driven Decision Making: Student Affairs										
.79 Assessment of SA Programs										
.77 Info. Dissemination: To High School Counselors	1									\uparrow
.72 Info. Dissemination: To Families	1									
.70 SA Staffing Policies			^							^
.67 SA Administrative Coordination										
Academic Affairs (AA) Scales										
.90 Data Driven Decision Making: Academic Affairs										
.85 FY Seminars	1			\uparrow				$\mathbf{h}\mathbf{h}$		
.84 Senior Faculty Teaching FY Courses		1							1	
.82 Recent Assessment Efforts: Persistence										
.81 Recent Assessment Efforts: Learning										
.73 Faculty Involvement Considered in Personnel Decisions				^						^
.71 Early Alert/Intervention Initiatives		^								ተተ
.69 Funding for Professional Development			^	ተተ	\uparrow	^	\uparrow		\uparrow	ተተ
.69 Curricular Emphasis on Diversity	^							$\downarrow \downarrow \downarrow$		
.65 Campus-Wide Admin. Coordination										
.65 Faculty Participation in FY Events										

¹NSSE Benchmarks are: EEE (Enriching Educational Experiences); SCE (Supportive Campus Environments); AC (Academic Challenge); SFI (Student-Faculty Interaction); ACL (Active & Collaborative Learning) ² Deep Learning indicators are: DPv1 (calculated using 12 individual items); DPv2 (calculated as the average of the 3 deep learning sub-scales: Higher Order Thinking, Integrative Learning, & Reflective Learning)

³ "Gains In" scales are: PRC (Practical Competence); PSO (Personal & Social Development); GED (General Education)