

Developing an Instrument to Observe and Evaluate Assessment System Maturity

Royce L. Robertson
Le Moyne College

Martha J. Larkin
Walden University

Maturity models fill the need for a measure of overall growth in how outcomes assessment is defined and how feedback informs improvement to meet internal and external expectations at individual institutions of higher education. An exploratory qualitative study was conducted to develop an instrument, the Assessment System Maturity Matrix (ASMM), to observe and evaluate outcomes assessment system maturity. Research procedures included conducting a literature review and comparing business, software, and limited education maturity models; administering self-reported questionnaires to higher education outcomes assessment professionals; and piloting the instrument. Data collection results provided the ideal criteria for the ASMM design and content and led to ASMM development and refinement. Some deviation from the established maturity order sequence of plan, build, implement, and evaluate was noted. The ASMM can enable diverse institutions of higher education to focus on current assessment practice maturity efficiently to determine subsequent actions to improve outcomes assessment performance. Further research is required to explore the awareness and development of outcomes assessment system maturity awareness among the various types of institutions within the higher education community.

Keywords: *higher education, assessment system, maturity, development*

Introduction

The term *maturity* refers to full development and is applicable to a variety of disciplines. Maturity models and systems are present in industries parallel to and overlapping with higher education such as business and technology. The key elements of most maturity models include defined levels of progression and criteria for evaluating levels. Institutions of higher education regularly seek new ways of completing routine tasks as an area for growth. As more emphasis is placed on assessment and accountability in higher education and as technological advances enable opportunities for more complex and robust assessment systems, a need exists to define and measure assessment system maturity. This can help institutions to determine where they are on the path to developing a mature assessment system and can provide a process for developing maturity to assist in planning and evaluating assessment goals. Defining and measuring assessment system maturity must consider numerous variables that can differ greatly among institutions (e.g., institution size and program focus, assessment goals, resources allocated to assessment, etc.). Drawing from literature in a variety of disciplines as well as a questionnaire constructed explicitly for the study, the researchers developed and piloted the Assessment System Maturity Matrix (ASMM, see Appendix A) to foster self-study and thoughtful observation of assessment practices to stimulate more intrinsically motivated improvement in assessment practices at American institutions of higher education.

Literature Review

An important distinction exists between assessment systems and plans with each having a specific purpose in the observation and evaluation of outcomes assessment and maturity in higher education. Those systems and plans are influenced distinctly by longstanding cultures and traditions at each institution, not to mention recent external pressures from regulatory agencies. These phenomena create conceptual incongruities between assessment for continuous improvement and assessment for compliance paradigms. Ewell (2009) emphasized that institutions typically adopt a stance reflecting both paradigms. The results of examining existing maturity models may assist to strengthen the relationship between continuous improvement and compliance. The following literature review provides a summary of assessment systems and plans as well as maturity models that are described in the context of higher education examples. Ewell (2002) noted that the brief history of higher education assessment depicts that models, systems, and techniques are critical to an institution's assessment development. He contended that systems thinking found quick application in higher education because of emphasis on scientific management. By design, systems thinking identifies various metrics and voices within the organization. Both quantitative and qualitative measures can be used to define systems and develop models. Beyond a model itself, the definition, instruments, and implementation can serve as significant challenges if not managed properly. The management of these challenges has regularly been described in an assessment plan (p. 13).

Assessment Systems and Plans

The difference between assessment systems and plans can be refined to a few simple indicators (see Table 1). An assessment system describes organized principles or methods for what is intended to be accomplished through outcomes assessment. The assessment plan is a proposal to formalize annual assessment requirements (Ewell, 1988; Ewell & Jones, 1986). For example, the system may emphasize timely and frequent data collection; the plan specifies the exact data and schedule. The system also is a description of environment and resources surrounding assessment; whereas, the plan describes the cyclical use of resources. For example, the system may describe an environment where faculty are encouraged to use assessment data as part of their own scholarship, and the plan describes the series of specific workshops used to strengthen data use. Likewise, the system reflects and influences the assessment culture on campus; the plan describes a series of actions that reinforce and strengthen the culture. For example, if the system emphasizes transparency and open sharing of data like the National Institute for Learning Outcomes Assessment (NILOA, 2011) Transparency Framework, then the plan specifies actions to improve reporting via a series of websites or dashboards (Aiman & Davis, 2014). Negotiating the use of systems and plans can enable more optimized implementation of outcomes assessment; however, internal and external forces influencing the institution must be considered.

Table 1. *Comparison of Systems and Plans*

| System | Plan |
|---|---|
| Describes organized principles or methods for what is intended to be accomplished through outcomes assessment | Describes annual assessment requirements (Ewell, 1988; Ewell & Jones, 1986) |
| Describes environment and resources surrounding assessment | Describes the cyclical use of resources |
| Reflects and influences the assessment culture on campus | Describes a series of actions that reinforce and strengthen the culture |

As outcomes assessment permeates the culture of higher education, institutions must recognize influencing forces and respond accordingly. Kuh and colleagues (2015) stated that increased assessment activity in higher education may be driven by external demands of governments, accreditors, and others. Although such efforts to address accountability are sincere, something is missing. For assessment work to be effective, it must be "...driven by the needs and priorities of the institution and used by the institution itself" (p. 220). It can be challenging for institutions to focus their assessment work on answering pertinent internal questions, providing evidence of system implementation, and planning while simultaneously meeting external expectations. Ewell (2002) remarked that assessment can mean different things to different institutions. Kuh and colleagues (2015) also noted that due to the diverse nature of institutions, assessment cannot be a "one-size-fits-all approach" (p. 222). Therefore, it is important for institutions to explore, discern, and then implement systems and plans that are responsive to the unique institutional culture and needs.

Assessing the needs and culture demonstrated in assessment systems and plans takes on many forms in higher education. There are instruments used to assess operational aspects of the assessment process (Braskamp & Enberg, 2014; Ferrera, 2018; State University System of New York, n.d.) that resemble checklists as well as rubrics. Each instrument is focused on assessing different aspects of higher education assessment systems but no one assesses maturity specifically.

Having a system and plan is not enough; rather, there must be a means of observing and evaluating overall growth or outcomes assessment maturity. In a traditional sense, maturity implies developing, evolving, or perfecting. Regardless of whether an institution views assessment as an internal (improvement) or an external (accreditation) paradigm (Ewell, 2008; Gaston, 2018), maturity is a valuable indicator of success. Success, in terms of maturity, is based on the degree that the systems and plans are defined and described (formalized) and how feedback informs improvement (optimized). Maturity, which can be an ambiguous concept, is made observable using a maturity model. Maturity models fill the need for a measure of overall growth in how outcomes assessment is formalized and optimized at the institution.

Maturity Models

In general, maturity models have two elements: formalize and optimize. The purpose of formalizing is to define processes related to the development of a product or program. Formalization entails four phases: Plan, Build, Implement, and Evaluate (PBIE). Planning is intentionally determining actions to achieve an outcome. Building is constructing and testing all the components of a product or program. Implementing is executing the steps of the plan according to how the product or program was built. Evaluating is the act of reflecting on the prior three elements to make improvements to the overall product, program, or process used. Optimization involves accounting for the frequency and purpose of performing a task. There are five levels: initial, repeatable, defined, capable, and efficient. Initial (ad hoc) is the starting point for a new task or process for which no previous documentation exists. Repeatable is performing a specified task the same way each time when there is sufficient previous documentation. Defined is when a specified task becomes standard process. Capable occurs when groups agree on the metrics used to manage the process and measure success. Efficient occurs when processes deliberately are improved based on the metrics used.

Maturity models, commonly found in business and software development circles, are used to observe formality and optimization of systems employed to produce the desired result. For example, if a software development company desires to improve how the application versions are developed and released, then a maturity model is used to formalize and optimize the process. Likewise, in higher education, maturity models can be used to formalize and optimize how assessment systems and plans are implemented. Several maturity models were analyzed to develop the ASMM first versions (see Appendix B). These models, because they are generally conceptual or theoretical, have not been

translated into operational checklists or rubrics used to assess what each of the referenced models purport. The significance of the ASMM research is the development of an operational instrument that reflects aspects of various models. There appears to be no one maturity model that is most popular or widely applicable because each one is designed to address specific disciplines.

Humphrey's (1987) research pioneered the capability maturity model (CMM) to describe software development process in five phases: initial, repeatable, defined, capable, and efficient. The two phases most critical to establishing maturity are defined and efficient. Being defined demonstrates a degree of formality, determined by Humphrey to be an observable trait of process maturity. Likewise, being efficient exhibits a level of optimization, determined by Humphrey as a critical component of organizational growth. He indicated that maturity occurs when routinized tasks are recognized, described, and performed continuously until improvement is evident. The CMM principles of formality and optimization can be applied to higher education. Outcomes assessment requires formality to define what assessment processes and actions will be performed within in each timeframe. Likewise, as institutions grapple with difficult organizational changes, optimization can be applied to ensure outcomes assessment is accomplished as efficiently. Maturity, based on the CMM, occurs when actions are employed that allow the institution to move its outcomes assessment process from one that is solely defined to one that has been optimized through revision. The work of Humphrey informed the design and content of the ASMM in the following ways: (a) viewing assessment system maturity as a process translates to observable actions that become the criteria (rows) of the instrument, (b) the emphasis on routinized tasks is represented in the actions and language used in the instrument, and (c) the five phases became the units of performance in the instrument, with efficient (optimization) being the ideal goal.

The CMM's influence had significant impact on the software development industry for more than a decade. Its use beyond software development was not immediate, particularly in more traditional settings such as institutions of higher education and their outcomes assessment processes. Kendra and Taplin (2004) explored maturity models using the perspective of institutional variables that influence maturation through the change agent competencies model. The researchers indicated that specific maturity model aspects included: project manager skills and competencies, performance measurement systems, organizational structures at the project level, and supporting management practices. The presence of managers and managerial skills (e.g., performing needs analysis, managing tasks and deadlines, applying governance for decision making, etc.) reflects the need for a maturity model to contain observable descriptors of how outcomes assessment is performed throughout the institution. The design and content of the ASMM was influenced by Kendra and Taplin in the following ways: (a) The importance of project management skills and competencies translates to how outcomes assessment is implemented; the presence of clear, repeatable plans makes operations more efficient, therefore, more mature. (b) The ASMM is designed to translate organizational structures of the institution into how maturity is observed at the course, program, department, and institutional levels.

The study performed by Hammer (2009) produced the process and enterprise maturity model (PEMM), which articulated five areas related to maturity: design, performers, owner, infrastructure, and metrics. PEMM was predicated on models related to business and industry; however, its applicability to higher education is clear given that institutions often have processes and products that must be managed. The content of PEMM describes compatible roles and assets in outcomes assessment system maturity. For example, the performers are the faculty and individuals responsible for fulfilling assessment requirements, whereas the owner is likely a director, dean, or provost depending on the size of institution. These areas can be aligned with and are reflective of common factors in success of an outcomes assessment initiative. The design and content of the ASMM was influenced by Hammer in the following way: defining actions based on roles promotes an

instrument that reflects the various levels of the institution and how each influences assessment maturity.

A model developed by Shepherd (2009) appears to be the first maturity model specific to higher education outcomes assessment; thus, it should be included as foundational to any other higher education outcomes assessment maturity models. He stressed the importance of systems moving from a less ordered, less valuable state to a well-organized, well-structured state. The levels of Shepherd's model include (a) three assessment areas of development, delivery, and presenting results; (b) four phases of ad hoc, managed, refined, and aligned; and (c) six measures of stakeholder satisfaction, assessment security, strategic alignment, predictable processes, trustworthy data, and communications with stakeholders.

Shepherd's model reflects an emphasis on stakeholders and two facets (satisfaction and communication). Both facets are critical to assessment system maturity because faculty (as the stewards of assessment) can withhold their support of assessment initiatives. Clear, consistent, timely communication can increase such support. This reflects an opportunity for various roles of academic structure (i.e., faculty, chairs, deans, etc.) to observe, document and improve assessments used with the system. The design and content of the ASMM was influenced by Shepard in the following ways: (a) Each criterion in the ASMM was designed from the less ordered and less structured to the more ordered and more structured, and (b) each criterion in the ASMM is written to have observable, documented indicators of maturity.

Powers (2011) depicted the maturity of web content to exhibit the following design themes: context, channel, consistency, community, consumer (behavior), and criterion. The element of context is critical to higher education given a need for transparency of information, especially those data related to achievement of student learning outcomes as a matter of public trust. Given the importance of recognizing context, it is necessary to analyze the audience consuming the content as well as its need and purpose. Maturity is achieved through differentiating content by audience with continual revision to improve its accuracy and relevance. For example, an immature outcomes assessment organization may rely entirely on aggregated data reports for consumption by large and small campus constituencies to discuss student learning outcomes. Alternately, a more mature organization will disaggregate the content of student learning outcome assessment reports as a means of engaging specific constituencies. Overall, the Powers model is most applicable to the ASMM in terms of communicating the data, not necessarily the process of collecting the data. The ASMM, informed by Powers, documents the processes related to how an institution matures its communication (e.g., the process) of outcomes assessment, not specifically what is communicated (e.g., the data or information).

The maturity model for information and communications technology in the school environment (ICTE-MM; Solar, Sabattin, & Parada, 2013) consists of two key domain areas: information criteria and information technology resources. Like other models of software and business development maturity (Hammer, 2009; Humphrey, 1987), this model provides direct connection to assessment system maturity in ways specifically related to the quality of information (information criteria) as well as the technological (information technology resources) foundations of a mature system. The key domain areas (e.g., quality of information and technological foundations) then are assessed using an instrument with a scale of 1 to 5, ranging from *initial* to *optimized*. The ICTE-MM considers the multiple perspectives of stakeholders as they interact with quality of information and technological foundations. For example, this implies that simply writing rubric criteria or test items naturally leads to maturity. It does not. The ICTE-MM model informed the development of the ASMM because it assumes that the organization must evaluate current assessment instruments and the technology used to collect data on a determined cycle. The ASMM also anticipates that the process may be viewed differently by deans and provosts as compared to faculty members depending on any

operational goals related to more efficient data collection, and so on. Regardless, the ASMM indicates the observable actions necessary to mature the assessment system by promoting sustainable, optimized processes related to quality of information and technological foundations.

Marchewka (2013) applied Humphrey's (1987) work to the accreditation of business schools, resulting in a five-level model, including initial, repeatable, defined, managed, and optimized, which described the core actions taken that move the practices from lower to higher maturity. *Assurance of learning* (AoL), a term widely used in colleges of business, focuses on the "continuous improvement of curriculum development, program review, and, in many cases, accreditation" (Marchewka, 2013, p. 2). This may include a range of core actions including: defining goals, determining actions, reviewing existing processes, communication between groups, complying with process steps, training or development, and process revision. Marchewka assumed that AoL (or outcomes assessment) cannot improve without first having competent individuals designing and implementing the assessment. For example, moving from initial to repeatable, Marchewka contended, depends on an individual's competence. To move from repeatable to defined, processes must be standard and consistent. To move from defined to managed, processes must be predictable for the individual. To move from managed to optimized, the individual must be engaged in continuous improvement by taking actions based on what is learned while implementing the assessment system. The implication is that individuals who design and implement the assessment system must look beyond the simple compliance of completing an action. Maturity is transactional and has many moving parts. The AoL model influenced the development of the ASMM by not treating the instrument criteria as a simple checklist of completed tasks. The ASMM is designed to treat outcomes assessment as interplay between compliance and continuous improvement, which influences how assessment is implemented at the institution, as also noted in Zilora & Leone (2014).

Aiman and Davis (2014), much like Humphrey and Marchewka, described a maturity model consisting of progression from ad hoc to optimized; however, the model also accommodates maturity at organizational levels and purposes including individual, department, enterprise, optimize, and innovate. When considering the three levels in combination, adoption will be more widely assumed and maturity will be more observable. The model supposes that optimization and innovation cannot occur fully without the achievement of mature practices, and those teams should be promoted by the organization as "centers of excellence." For example, if the mathematics department demonstrates excellence in valid and reliable test item construction, then it should serve as a center to assist other departments, raising the overall institutional maturity level. The center of excellence concept illuminates niche processes that are focused on the individuals' competence in terms of how they influence work at the department level. The ASMM reflects Aiman and Davis in the following ways: (a) The ASMM is intentionally designed to document assessment work performed at department level first, then departmental results can be aggregated to demonstrate the overall innovation and optimization within the organization. (b) The results of departmental use of the ASMM, aggregated into an institutional profile, can be used to stimulate "act local, results global" impact. (c) The ASMM is designed to identify matured processes that may serve as centers of assessment excellence to develop, test, refine, and disseminate various assessment practices throughout the institution.

Studying and summarizing maturity models strengthens the relationship between compliance and improvement. Consistent throughout the models are key themes related to compliance through formalization and improvement through optimization. If the institution engages in formalizing outcomes assessment processes, then routine compliance tasks become second nature. Also, when the institution wants improvement as the desired goal of implementing outcomes assessment, then it can be achieved through efforts to optimize how outcomes assessment is performed at the institution. Essentially, as stated by the NILOA (2016), "Focus on improvement and compliance will take care of itself" (p. 6). However, these improvements cannot depend solely on the institution's will. There must

be a combination of leadership, individual skill, technology, communication, and collaboration applied for the desire to improve to be translated to the capacity to optimize.

Methodology

A thorough literature search did not provide a comprehensive instrument for measuring assessment system maturity in higher education that can be tailored to goals and needs of diverse types of institutions. Therefore, the researchers engaged in a study to address such a need. The ongoing, developmental, qualitative study used a combination of Type 1 (i.e., instrument development, evaluation, and validation) and Type 2 (i.e., model development, use, and validation) development research (Richey & Klein, 2005) and included a variety of steps, ranging from thorough and extensive literature review as well as opportunities for the professional community to provide feedback on the instrument. The instrument, organized like the PEMM (Hammer, 2007), consists of global categorical areas such as leadership, governance, infrastructure, and accountability, and so on, which were drawn from related literature. To minimize confusion, the study's purpose was not to dissect nor evaluate assessment plans; rather, to develop a model for observing assessment system maturity. An assessment plan emphasizes action and implementation. Observing assessment system maturity emphasizes formality and optimization of implementation of all aspects of outcomes assessment. The literature review alone does not serve as sufficient evidence for the development of the entire instrument; therefore, a purposeful, knowledgeable sample population would provide the expertise to design an instrument necessary to measure assessment system maturity.

The study used a convenience sample to solicit feedback. The Assessment Institute, hosted by Indiana University–Purdue University, Indianapolis, is the premier American higher education outcomes assessment conference. Because of their distinct roles (Nicholas & Slotnick, 2018), attendees and presenters are considered thought leaders in outcomes assessment. Therefore, input from this population could be valuable in the creation of resources used to improve assessment practice. Participants (i.e., conference attendees) possess explicit professional knowledge and skill related to the phenomenon being studied. For example, participants who work directly in outcomes assessments are highly likely to engage in daily work related to assessment planning and evaluating the institution's assessment system.

The sample size is reflective of qualitative research, which emphasizes understanding the lived experience. In qualitative research, however, the sample size generally is not predetermined. Given that the study aimed to develop an instrument, the number of participants adequately informed all important elements related to the development of the instruments. In general, the participants had substantial years of experience and significant role with within their institutions to determine outcomes assessment-related phenomenon used in the development of the instrument. As further development of the instrument occurs, new participants will be solicited that hold equal experience and stature in their institutions. The participants in both instrument revision cycles work directly with outcomes assessment at their institution. In the first instrument revision cycle, seven of 10 position titles included assessment, while one also included institutional research and one was an associate dean. In the second instrument revision cycle, four of five titles were associated with jobs in assessment, while the fifth was a faculty member.

Table 2. Instrument Revision Cycles Demographic Profile of Participants

| | <i>n</i> | Cycle 1, % | Cycle 2, % |
|--|----------|------------|------------|
| Type of institution | | | |
| Community College | 1 | 9 | 0 |
| Public | 6 | 55 | 60 |
| Private Religious | 1 | 9 | 20 |
| Private Secular | 2 | 18 | 0 |
| Private Proprietary | 1 | 9 | 20 |
| Highest degree awarded | | | |
| Associate | 1 | 9 | 0 |
| Bachelor | 1 | 9 | 0 |
| Master | 4 | 36 | 40 |
| Specialist | 0 | 0 | 0 |
| Doctorate | 5 | 46 | 60 |
| Postdoctorate | 0 | 0 | 0 |
| Number of full- and part-time students | | | |
| ≤5,000 | 6 | 54 | 40 |
| 5,001–10,1000 | 1 | 9 | 0 |
| 10,001–20,000 | 0 | 0 | 0 |
| 20,001–25,000 | 2 | 18 | 40 |
| 25,001–30,000 | 0 | 0 | 0 |
| ≥30,001 | 2 | 18 | 20 |
| Regional accreditor for institution | | | |
| New England | 1 | 9 | 0 |
| Middle States | 1 | 9 | 0 |
| Southern | 2 | 18 | 20 |
| North Central | 5 | 46 | 60 |
| Northwest | 2 | 18 | 0 |
| Western | 0 | 0 | 0 |

Table 2 shows demographic distribution of participants according to institutional type, highest degree awarded by the institution, number of full- and part-time students, and regional accrediting body. The demographic data from Table 2 were important when considering how a participants' response may be influenced by the type, size, or region of the institution.

From literature and document reviews, the first instrument draft was developed and shared with the population sample for initial feedback. Their feedback was used to identify specific design and content improvements to the instrument. The second and third drafts were shared with the participants and the feedback and revision processes was repeated. Data and resulting analysis were included in presentations to various constituents of higher education outcomes assessment professional community. Continued implementation and analysis will be used to further refine the ASMM.

Findings

Through the implementation of the methodology, the researchers collected data associated with three described cycles. Each cycle produced both self-reported questionnaire and narrative feedback that was used to revise instrument design and content.

First Instrument Revision Cycle

The purpose of the first instrument revision cycle was to solicit feedback on the design and content of the ASMM's first draft using a two-part questionnaire. The first part solicited demographic data. The second asked participants to prioritize various dimensions of outcomes assessment work in higher education. Assuming that not all participants had knowledge of maturity models, a basic definition of maturity in context to outcomes assessment was provided: a process of development that every institution engages to respond appropriately to the needs presented by participating in outcomes assessment. Then, participants were provided instructions to review the list of thematic areas, prioritizing them from lowest to highest without repeating any rankings. Once the prioritizing was completed, participants were asked to review a sample criterion (row) of the ASMM and provide an open-ended response regarding the design and content of the row.

Table 3. *Thematic Areas Prioritized by Participants in First Instrument Revision Cycle*

| Rank Level | Rank Number | Weighted Mean | Thematic Area | Maturity Area |
|------------|-------------|---------------|---|---------------|
| High | 1 | 6.22 | Outcomes assessment as continuous improvement | Evaluate |
| | 2 | 7.71 | Data analysis | Implement |
| | 3 | 9.11 | Data collection | Implement |
| | 4 | 9.29 | Faculty involvement | Build |
| | 5 | 9.43 | Presence of an assessment team | Plan |
| Middle | 6 | 9.71 | Culture toward assessment | Build |
| | 7 | 12.29 | Value of engagement in assessment | Build |
| | 8 | 12.43 | Planning for assessment | Plan |
| | 9 | 13.86 | Use of both direct and indirect evidence | Plan |
| | 10 | 13.38 | Outcomes development at the institutional level | Plan |
| Low | 11 | 14.00 | Budgeting for assessment | Plan |
| | 12 | 14.25 | Professional development on assessment | Implement |
| | 13 | 17.14 | Technology for assessment | Build |
| | 14 | 19.33 | Outcomes assessment for accreditation | Implement |
| | 15 | 19.67 | Recognition for participation in assessment | Build |

Table 3 depicts data collected from participants prioritizing a variety of thematic areas related to outcomes assessment. A weighted mean was assigned based on how participants ranked each thematic area. The weighted average was calculated based on each theme to ensure that a mean of low and high rankings was not treated the same as a consistent set of rankings.

Discrepancies between the prioritization (rank number) and the order of maturity areas may illuminate differences between what participants determine is important as opposed to what is mature. While the established order of maturity areas is *plan*, *build*, *implement*, and *evaluate*, the prioritization (rank number) by participants did not reflect the traditional PBIE sequence. Likewise, considering that the plan phase is the most foundational aspect of developing system maturity, the participants indicated that items in higher phases took priority than those in the plan phase. For example, the thematic area with the highest rank number is outcomes assessment as continuous improvement, which is part of the maturity area titled evaluate (the highest level). Alternately, most foundational plan-related thematic areas have a middle ranking. The prioritization data also revealed some inconsistencies in logical progression of PBIE development. For example, data analysis was prioritized one placeholder higher than data collection, if in fact data should be collected before it is analyzed. Also, data collection was prioritized significantly higher than use of both direct and indirect evidence, which is a prerequisite of selecting sources of data. Finally, faculty

involvement—which is commonly cited factor in outcomes assessment success (Hutchings, 2010)—is ranked significantly higher than recognition for participation in assessment.

Participants ranked the definitions and phases of maturity according to different priorities. Institutions of higher education often desire end results such as achievement of continuous improvement, a faculty body engaged in assessment, or the act of data collection and analysis. However, all those items, deemed as high priority by the participants, predominantly are antecedent outcomes of clear planning and building. Planning and building are considered foundational aspects of developing maturity yet have been prioritized lower by the participants of this study. The tangible actions (e.g., collecting and analyzing data) may be considered a higher priority than developing the foundations of assessment system maturity because of many institutional reasons. For example, identifying faculty involvement as a high priority may be the result of changes in institutional culture or a decrease in human and fiscal resources. These phenomena may produce questions for further research regarding the rationale for prioritization and the instructional culture and history influencing prioritization. In terms of the ASMM development, the instrument is designed to prioritize the foundational elements of maturity first. Specifically, planning comes before building which is before implementation and then ending with evaluation.

Prioritizing outcomes assessment action also produced some contradictions. First, the representation of continuous improvement and accreditation raised questions of polarity. The highest priority for participants was outcomes assessment as continuous improvement (an evaluate item). Likewise, prioritized outcomes assessment for accreditation is categorized (an implement item) as second to last (14 of 15) by the participants. This polarity highlights the internal–external struggles institutions experience between continuous improvement and accreditation. Each item is the result of thoughtful plan and build phases. The ideal scenario might indicate that formality and optimization of assessment systems can engage continuous improvement while simultaneously achieve accreditation success. The polarity seems to illustrate the difference between priority and maturity. As a result, the ASMM was designed to show a relationship between accreditation and continuous improvement as related to the maturation of an assessment system.

Table 4. *Distribution of Participant Agreement With Assessment System Maturity Matrix Design*

| Design (D) Items | Total | | Strongly Disagree | | Disagree | | Neither Agree or Disagree | | Agree | | Strongly Agree | |
|--------------------------------------|----------|-----|-------------------|------|----------|------|---------------------------|------|----------|------|----------------|------|
| | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % |
| D1: Four performance levels | 11 | 100 | 0 | 0 | 2 | 18.2 | 1 | 9.1 | 5 | 45.5 | 3 | 27.3 |
| D2: Labeling of criteria | 11 | 100 | 0 | 0 | 2 | 18.2 | 1 | 9.1 | 7 | 63.6 | 1 | 9.1 |
| D3: Use of essential questions | 11 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 63.6 | 4 | 36.4 |
| D4: Lowest to highest, left to right | 11 | 100 | 2 | 18.2 | 0 | 0 | 2 | 18.2 | 3 | 27.3 | 4 | 36.4 |

Note. Exact wording of design items are as follows—D1: Using four performance levels (i.e., the number of columns) can effectively accommodate varying observations; D2: The labeling of criteria (e.g., Assessment Maturity Matrix Implement 1.1) is appropriate for succinctly identifying specific rows; D3: The use of a guiding question in the Elements/Essential Question column (Appendix A) is helpful for directing the purpose of criteria (e.g., “At what levels of the institution do outcomes exist?”); D4: The progression of lowest to highest from left to right is more appropriate than highest to lowest.

Table 4 depicts participants' agreement with various aspects of the ASMM design. Common design items included performance levels (e.g., Item D1) or order of performance levels (e.g., Item D4). The narrative feedback on the columns' design appeared to be influenced by personal preference as well as institution type. The feedback indicated that performance descriptors in Item D1—start, low, or high—needed revision. One participant mentioned that the four-level distinction in Item D1 (course, department, college, institution) does not apply to their institution while another stated that inclusion of course, program, and institutional outcomes were not applicable. Also, some participants indicated that terms used in Item D2 to describe each level must be clear and incremental for a Likert instrument like the ASMM. Therefore, design should be adjusted to reflect varying institutional types. In Item D3, all participants either agreed or strongly agreed that use of essential questions was critical to the ASMM design. This indicates that use of essential questions strengthens clarity and focus of each instrument criteria. In Item D4, some favoring a column order of highest to lowest indicated that they prefer “that the aspirational level appears first.” Some stated that their responses would be further influenced by reviewing the full rubric. Translating these results to the development of the ASMM, three of the four features (i.e., D1: Four Performance Levels, D2: Labeling of Criteria, D3: Use of Essential Questions) included were agreed upon by the participants. One feature (i.e., D4: Lowest to Highest, Left to Right) was used (despite mixed opinion) because of the progressive nature of maturity beginning with the least ideal state.

Table 5. Distribution of Participant Agreement With Assessment System Maturity Matrix Content

| Content Items | Total | | Strongly Disagree | | Disagree | | Neither Agree or Disagree | | Agree | | Strongly Agree | |
|--------------------------------------|----------|-----|-------------------|---|----------|------|---------------------------|------|----------|------|----------------|------|
| | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % |
| C1: Titles | 11 | 100 | 0 | 0 | 4 | 36.4 | 1 | 9.1 | 5 | 45.5 | 1 | 9.1 |
| C2: Essential question content | 11 | 100 | 0 | 0 | 1 | 9.1 | 2 | 18.2 | 6 | 54.5 | 1 | 9.1 |
| C3: Appropriate institutional levels | 11 | 100 | 0 | 0 | 4 | 36.4 | 2 | 18.2 | 3 | 27.3 | 2 | 18.2 |

Note. Exact wording of content items are as follows—C1: The titles of levels (i.e., start, low, medium, high) are appropriate for adequately describing the levels; C2: The content of C2 is properly reflected in the performance criteria provided in each of four columns (e.g., “Outcomes are written for a specific segment of the institution...” in the Start column properly relates to C2, “At what levels of the institution do outcomes exist?”); and C3: The content of each of four cells (e.g., start through high) adequately reflect the continuum of performance exhibited by an institution related to C2. For example, “Outcomes are written for courses, programs, departments, schools/colleges, and the institution” is appropriate for the high level of maturity.

Table 5 depicts participant's agreement with ASMM content. Common content items included criterion titles (e.g., the criteria in the left-most rubric column) or institutional organizational levels (e.g., department, school, college, institution). Narrative feedback column content also appeared to be influenced by personal preference. For Item C1, participant comments illustrated general disagreement with titles of levels in the sample. The combination of start, low, medium, high was the title range provided. However, titles used were adapted from literature related to maturity systems developed for assessment of learning (Marchewka, 2013). Some participants' suggestions included more clearly differentiating start and low levels. Also, some participants preferred using numbers only without titles. Other participants preferred a different set of titles such as beginning, developing, acceptable, and exemplary. In Item C2, while the participants predominately agreed or strongly agreed, there were no narrative comments to analyze. In Item C3, one participant preferred language in each criterion that qualitatively describes maturity rather than language that simply

counts the indicators of maturity. Translating these results to the development of the ASMM, two of the three features (i.e., C1: Titles, C2: Essential question content) included were agreed upon by the participants. One feature (i.e., C3: Appropriate institutional levels) was adjusted to fit the feedback from the participants, reflecting a more streamlined organization for the institutions.

Second Instrument Revision Cycle

The purpose of the second instrument revision cycle was to solicit feedback on the design and content of specific criteria (rows) of the ASMM using a combination of open-ended prompts and Likert-type scale items. The Likert-type scale items focused the participants' agreement with the content of specific criteria as related to outcomes assessment system maturity. The open-ended questions asked participants to provide feedback on any criteria not included as well as any revisions to language of the criteria.

Opinion differed as to the importance of planning as a critical stage in establishing assessment system maturity. Two participants indicated it was Most important while two indicated it was least important. One participant who supported the importance of planning indicated, "Planning sets the stage for the effectiveness of the system. The planning team must include representation of the entities building, managing, and evaluating the effectiveness of the system." Another participant indicated,

Proper planning and management of assessment system/resources can reduce the need for building much of a formal physical or electronic structures [sic]. It may be chaotic but having a plan and method for navigating the chaos can override placing importance on building something.

These comments reinforced the ASMM design element of plan being the initial phase in the maturity of the assessment system. Those ranking planning as less important cited that evaluation is important but cannot occur until planning as well as building and managing have occurred. This is contradictory to all the maturity models cited because planning is always considered a prerequisite for building, managing and evaluating. In support of this idea, one participant stated, "your system is not really mature until you are at the stage of evaluating the system which only happens after it has been planned, built, and managed." Per many of the participants, the actions of building the assessment system appeared to be of lesser importance in establishing maturity than the other areas (i.e., planning, managing, and evaluating). Four of five participants indicated that planning was not important or least important. This inconsistency may be related to differences in definitions between the models used and the participants' perspectives. According to the participants, managing the assessment system appeared to be less important when establishing maturity than compared to survey item choices. Four of five participants indicated that managing was important. According to participants, evaluating the assessment system appeared to be of great importance in establishing maturity than the other areas. Four of five participants indicated that evaluating the assessment system was most important. Overall, though, the participants' consensus indicated that all components are necessary for maturity; however, evaluating the assessment system appeared to be the most important in terms of actual survey responses. While these results are inconsistent with the design of the ASMM, the design explicitly reflects common maturity models. Using the participant results to change the maturity model as an influence reduces its validity.

Considerations Specific to Design

General agreement was observed for organization of the matrix into broad categories of PBIE that effectively reflects phases used to establish maturity. Three of five participants indicated agreement with the organization, while two indicated neither agreement nor disagreement. General agreement

was observed for the use of the labeling scheme (e.g., P1, B1, I1, E1) effectively organizes the criteria for measuring maturity (i.e., the rubric rows). Four of five participants indicated agreement with the labeling scheme while one indicated neither agreement nor disagreement. There was considerable agreement for including the not applicable (N/A) choice for each criterion to improve instrument design allowing for differentiation between the lowest levels (e.g., start or low) and those criteria not applicable to the institution. Four of five participants indicated agreement with addition of N/A for not applicable designations while one indicated strong agreement. Considerable agreement was observed for including the no evidence (N/E) choice for each criterion. This improves instrument design by clearly identifying when no evidence is available to support the criteria score. Five of five participants indicated agreement with addition of N/E for no evidence designations. General agreement was observed for using bold to emphasize the keywords in each criterion as an effective design strategy. Three of five participants indicated agreement with use of bold while one indicated neither agreement nor disagreement and one indicated strong agreement. Considerable agreement was observed for revising performance levels (i.e., column headers) to represent the choice of getting started, progress in limited areas, progress with larger groups, consistency across campus. Four of five participants indicated agreement with this choice. One participant also selected beginning, developing, acceptable, and exemplary. All the items in this section were included in the design of the ASMM because of consensus with the participants' feedback.

Considerations Specific to Content

Differing opinion was observed for revising terminology used to represent description of organizational units within the institution. Two of five participants preferred the existing choice of "courses, programs, departments, schools/colleges, or the institution." Two of five participants preferred the alternative of does not exist, progress in limited areas, progress with larger group, and consistency across campus. One participant also indicated that the instrument "Does not make these hierarchical distinctions." Because of the lack of consensus to select other verbiage, the ASMM content remained the same as the original version.

Results indicate general agreement that content of each criterion appropriately represents internal forces which influence assessment. Four of five participants indicated agreement with importance of internal forces, while one indicated neither agreement nor disagreement. Results indicated general agreement with content of each criterion appropriately represents external forces that influence assessment. This led to maintaining the existing language of the ASMM. Four of five participants indicated agreement with the importance of external forces, while one indicated disagreement. Results indicated weak agreement with content of each criterion appropriately represents regulatory forces that influence assessment. This led to maintaining the existing language of the ASMM. Three of five participants indicated agreement with the importance of regulatory forces, while one indicated disagreement and one indicated neither agreement nor disagreement. Important insight was gleaned through the narrative comments. Numerous comments reflected support for evaluating assessment processes (often referred to as "assessing how we assess"). As one participant commented, "evaluating [our] system/process is important, too, as this can impact any future efforts to build or solidify what [we're] doing." Results also indicated general agreement with content of each essential question appropriately represents the content of the related criterion. This led to maintaining the existing language of the ASMM. Five of five participants indicated agreement with the use of essential questions. Results indicated general agreement with section titled plan containing a disproportionate number of criteria in comparison to build, implement, and evaluate. This led to maintaining the existing language of the ASMM. Three of five participants indicated neither agreement nor disagreement with the number of items in plan, while one participant indicated agreement and one indicated disagreement. Some participants appeared to believe that doing assessment seems more important than planning for assessment.

Third Instrument Revision Cycle (Pilot)

The purpose of the third (pilot) instrument revision cycle was to solicit feedback on the design and content of the entire ASMM because of the participants assessing their own institution using the instrument. A Likert-type scale item was used to solicit feedback on the content of all criteria as related to how the participant assessed their own institution. The open-ended questions asked participants to provide feedback on any criteria not included as well as any revisions to language of the criteria.

The demographic profile of participants in the pilot was the same as the second round of review and revision. The pilot outcome was to have participants observe and assess maturity of their own institution as means of identifying practical revisions to the instrument. Data resulting from the pilot was organized per the criteria (rows) and maturity level (columns) presented in the ASMM (see Appendix A). As a supplement, open-ended survey items were added to the end of each ASMM section to solicit design and content feedback from pilot participants.

To summarize pilot results of the Preliminary Considerations section, most items ranged from low to high with a majority at the medium level for many of criteria. For example, 57.4% of the pilot participants reported, “Outcomes are written for three specific segments of the institution, including either courses, programs, departments, schools/colleges, or the institution, but only three of these” at the medium level. Likewise, 100% of “Outcomes of various levels are shared on syllabi, reflected in course descriptions, and published on institutional webpages” and 86.7% of “Adequate, multiple measures are used to describe student performance” also reported at the medium level. The item with the greatest range was “How are those responsible for assessment-related work defined and organized?” The results revealed that equal proportion of pilot participants reflected low and medium levels, described as “Those responsible for assessment are defined using simple methods. Either one individual is assigned specific responsibilities, or multiple individuals assigned the smaller duties on a smaller scale” and “Those responsible for assessment are multiple individuals performing similar tasks in different areas, requiring high levels of coordination to organize the complexity of the interactions” respectively. Institutions may have varying degrees of foundational supports and systems necessary for assessment processes to mature.

When considering the plan section pilot results, most items ranged from low to high with a majority at the medium level for many criteria. For the criterion “How is the value of assessment established?” responses ranged from low (“Values are implicit, developed informally by a small collection of individuals, and are communicated in a limited fashion within the group”) to high (“Values are mission-driven, identified through reflection on the institution’s working culture. Appropriate stakeholders approve the values”). Also, for the criterion “How is planning for assessment done?” responses ranged from low (“The institution plans how it conducts assessment on a limited basis in more informal circles”) to high (“The institution plans how it conducts assessment through systematic activities and events”). Finally, for the criterion “At what levels of the institution does assessment most commonly occur?” responses ranged from low (“Assessment is used to meet individual, course, and program level needs”) to high (“Assessment is used to meet needs from individual to institutional”). Limited institutional planning for assessment and course and program focused assessment may impact maturity.

The build section pilot results indicated most items ranged from low to high with a majority at the medium level for many criteria. For the criterion “Who influences the culture of assessment?” responses ranged from start (“Temporary and/or individual groups influence or establish culture of assessment on an as needed basis [adhocracy]”) to high (“The institution establishes culture of assessment with proper influence from external professional groups [professional culture]”). Also, for the criterion “How are faculty and staff recognized for engagement in assessment activities?”

responses ranged from start (“The institution has no formal system for recognizing engagement in assessment activities where there are no specific types of recognition identified. There is no clear motivation to participate in assessment activities”) to medium (“The institution balances generic and strategic incentives for recognizing engagement in assessment activities where there is a menu of types of recognition used frequently. Motivation to participate in assessment activities tends to serve extrinsic expectations (i.e., accreditation) while informing intrinsic needs [i.e., course, program improvement]”). Finally, for the criterion “At what levels of the institution does assessment most commonly occur?” responses ranged from low (“Assessment is used to meet individual, course, and program level needs”) to high (“Assessment is used to meet needs from individual to institutional”). Institutions should examine how culture toward assessment and recognition for participation in assessment may be influencing overall assessment system maturity.

Review of the implement section pilot results revealed most items ranged from start to high with a majority at the medium level for many of criteria. For the criterion “How is the institutional definition of transparency determined?” responses ranged from start (“Individuals determine definition of transparency which may be limited in scope”) to high (“The institution adapts a definition of transparency and then revises it based on data collected”). Also, for the criterion “How is transparency managed?” responses ranged from start (“Faculty governance process does not discuss or minimally discusses transparency”) to medium (“Faculty governance process implements plans for addressing transparency through an individual committee”). Transparency of assessment system processes and data are defined loosely and fluid in implementation, and should be considered an emergent theme in maturity.

To summarize the evaluate section pilot results, most items ranged from start to high with a majority at the medium level for many criteria. For the criterion “To what degree are various actions and resources used to evaluate the process for approving outcomes?” responses ranged from start (“There have been no or minimal actions to recognize the strengths and weaknesses of how outcomes are approved. There are no resources or no commitments to evaluate approved outcomes”) to medium (“Reoccurring actions are planned to identify strengths and weaknesses of the process to approve outcomes across many campus areas. The institution makes adequate commitment of resources to evaluate approved outcomes”). Also, for the criterion “To what degree are various actions and resources used to evaluate transparency of assessment efforts?” responses ranged from start (“There have been no or minimal actions to recognize the strengths and weaknesses of how transparency in assessment is demonstrated. There are no resources or no commitments to evaluate transparency of assessment efforts”) to high (“Sustained actions are planned to identify strengths and weaknesses of how transparency in assessment is demonstrated across the entire campus. The institution makes effective commitment of resources to evaluate transparency of assessment efforts”). Evaluating existing assessment processes (e.g., the development and approval of outcomes) and conditions (e.g., the transparency of assessment processes and data) may be considered infrequently as a means of improving assessment system maturity.

When considering the four sections (i.e., PBIE), the results were consistent enough to maintain the overall intent of the criteria (rows) and the levels (column). Both the numeric results and any comments associated with the pilot use indicated little disagreement with content. In most cases, results were used to revise minor language conventions of each criterion to reduce ambiguity when observing assessment system elements.

Conclusion

The following actions should be considered to improve the ASSM in future revisions. First, the higher education outcomes assessment community should be made more aware of concepts in assessment system maturity. As described in the findings of the formative data collection cycles,

participants prioritized implementing and evaluating the system over planning and building the system. This is contradictory to assessment systems that reflect a framework where maturity requires planning and building as prerequisite conditions. To adequately recognize the breadth and depth of assessment system maturity, the professional community ideally would develop a shared language of foundational terms, concepts, and actions that reflect process milestones and criteria of common maturity models.

Second, the instrument design needs to accommodate more accurately the differences in how outcomes assessment systems are implemented at the department, program, or institutional level at different institutions. For example, a private, religious, liberal arts institution may approach outcomes assessment system differently than a private, secular, professional school. The intent, language, and logic of the ASMM achieve a cursory level of usability among various institutions; however, further piloting must occur to refine the instrument for maximum use at more institutions.

Third, the researchers need to solicit institutions to further pilot and norm the ASMM, and then make appropriate revisions to the instrument and the instructions describing its use. Piloting would involve creating and sharing a web-based interface where institutional representatives can follow specific directions to complete the instrument, and then provide feedback on the content and design. As a critical mass of institutions are added to the pilot, the demographic and maturity data may be used to identify areas of need to drive large-scale improvement regarding institutions perform outcomes assessment. The pilot phase of ASMM development produced promising initial results with clearly discernable areas for improvement. Feedback from future piloting cycles should be aggregated and analyzed to determine actions that will improve ASMM design and content.

Fourth, the researchers need to solicit multiple reviewers from a single institution to further pilot and norm the ASMM to determine its reliability, leading to appropriate revisions to the instrument and the instructions describing its use. When a critical mass of reviewers from one institution complete the ASMM, the results may be used to identify areas of need of programs within the institution where outcomes assessment processes can be matured. Feedback from institutional norming for reliability could be used to guide strategic planning for outcomes assessment.

Fifth, given that the researchers selected a defined maturity model without testing the participants' understanding of constructs related to the model. As a result, the response process validity is not established due to a lack of consistent understanding of the constructs between the researchers and participants. Future studies should include some type of process to ensure consistent interpretation of the instrument as well as information about decisions participants use when completing the instrument.

The pursuit of outcomes assessment maturity should not be treated as an add-on measure of success; rather, one integrated with the most essential components of continuous improvement. The most useful connection between maturity and outcomes assessment rests with the relationship between the evaluation phase and the optimization level. Evaluation is the concluding phase that results from planning, building, and implementing an outcomes assessment system. Likewise, optimization is the goal of maturity after processes are described, repeated, defined, and managed. Overall, if the desired result is to achieve the evaluate-optimized state, then any number of prescribed steps must occur. One complaint from the higher education community is that assessment appears to be a routine of compliance activities (Ewell, 2009). Compliance reflects plan, build, implement with some emphasis on evaluate as well as describe, repeat, define, and manage with some emphasis on optimize. If this is the case, maturity models and continuous improvement are similar in nature. Given that compliance can be considered a binary activity (e.g., complete or incomplete) and continuous improvement is a cyclic process of determining areas for improvement and brainstorming open-ended actions (e.g., refine and grow). NILOA (2016) advocated to "focus on improvement and

compliance will take care of itself” (p. 6). Although the sentiment is accurate, there are many detailed facets of changing how an institution approaches both compliance and continuous improvement. The ASMM serves as an instrument that can enable diverse institutions of higher education to determine efficiently areas for improvement by focusing on the maturity of current assessment practices and, thus, subsequent actions to improve how outcomes assessment is performed. The ASMM can also help institutions coalesce leadership, individual skill, technology, communication, and collaboration applied for the desire to improve to be translated to the capacity to optimize.

References

- Aiman, Z., & Davis, J. (2014). *Business transformation: A roadmap for maximizing organizational insight*. Hoboken, NJ: Wiley.
- Braskamp, L., & Engberg, M. (2014). *Guidelines for judging the effectiveness of assessing student learning*. Loyola University Chicago: Chicago IL.
- Ewell, P. T. (1988). Implementing assessment: Some organizational issues. *New Directions for Institutional Research*, 59, 15–28. doi:10.1002/ir.37019885904
- Ewell, P. (2002). An emerging scholarship: A brief history of assessment. In T. Banta (Ed.) *Building a scholarship of assessment* (pp. 3–25). Hoboken, NJ: Wiley.
- Ewell, P. (2008). Assessment and accountability in America today: Background and context. *New Directions for Institutional Research: Assessing and Accounting for Student Learning—Beyond the Spellings Commission*, 5, 7–17.
- Ewell, P. T. (2009). *Assessment, accountability, and improvement: Revisiting the tension* (NILOA Occasional Paper No. 1). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Institute for Learning Outcomes Assessment. Retrieved from <http://learningoutcomesassessment.org/OccasionalPapers.htm>
- Ewell, P. T., & Jones, D. P. (1986). The costs of assessment. In C. Adelman (Ed.), *Assessment in American higher education: Issues and contexts* (pp. 33–46). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Ferrera, V. (2018). *Assessing our assessment: findings and lessons learned three years later*. Proceedings of the 5th Annual Assessment Network of New York (ANNY) Conference, Syracuse, NY.
- Gaston, P. (2018). *Assessment and accreditation: An imperiled symbiosis* (NILOA Occasional Paper No. 33). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from <http://learningoutcomesassessment.org/OccasionalPapers.htm>
- Hammer, M. (2007). The process audit. *Harvard Business Review*. Retrieved from <http://hbr.org/2007/04/the-process-audit/ar/1>
- Humphrey, W. S. (1987). *Characterizing the software process: A maturity framework*. Proceedings of the Software Engineering Institute, Pittsburgh, PA.
- Humphrey, W. (1989). *Managing the software process*. Reading, MA: Addison-Wesley.
- Hutchings, P. (2010). *Opening doors to faculty involvement in assessment*. Retrieved from <http://www.learningoutcomeassessment.org/documents/PatHutchings.pdf>
- Kendra, K. & Taplin, L. (2004). Project success: A cultural framework. *Project Management Journal*, 35, 30–45.

- Kuh, G., Ikenberry, S., Jankowski, N., Cain, T., Ewell, P., Hutchings, P., & Kinzie, J. (2015). *Using evidence of student learning to improve higher education*. San Francisco, CA: Jossey-Bass.
- Marchewka, J. (2013). Applying the capability maturity model to assurance of learning. *Communications of the International Information Management Association, 13*, 1–8. Retrieved from <http://scholarworks.lib.csusb.edu/ciima/vol13/iss1/1>
- National Institute for Learning Outcomes Assessment (NILOA). (2011). *Transparency framework*. Retrieved from: <http://www.learningoutcomeassessment.org/TransparencyFramework.htm>
- National Institute for Learning Outcomes Assessment (NILOA). (2016). *Higher education quality: Why documenting learning matters*. Retrieved from http://www.learningoutcomesassessment.org/documents/NILOA_policy_statement.pdf
- Nicholas, M., & Slotnick, R. (2018, April). *A portrait of the assessment professional in the United States: Results from a national survey* (Occasional Paper No. 34). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from <http://learningoutcomesassessment.org/occasionalpaperthirtyfour.html>
- Powers, S. (2011). *The seven C's of web content management*. Retrieved from <http://www.percussion.com/resources/white-papers/forrester-research/index>
- Richey, R., & Klein, D. (2005). Development research methods: creating knowledge from instructional design and development practice. *Journal of Computing in Higher Education, 16*, 23–38.
- Shepherd, E. (2009). *Assessment maturity model*. Retrieved from <http://www.assessmentmaturitymodel.org>
- Solar, M., Sabattin, J., & Parada, V. (2013). A maturity model for assessing the use of ICT in school education. *Educational Technology & Society, 16*, 206–218.
- State University System of New York. (n.d.). *Assessing Institutional Effectiveness Self-Assessment*. Retrieved from <http://system.suny.edu/media/suny/content-assets/documents/academic-affairs/assessment/Institutional-Effectiveness-Rubric-Branded.pdf>
- Zilora, S., & Leone, J. (2014). *A mature approach to assessment*. Paper presented at ASEE 2014 Annual Conference and Exposition. Indianapolis, IN. Retrieved from <https://peer.asee.org>

[Appendix follows]

Appendix A

Assessment System Maturity Matrix

| Element/Essential Question | Maturity Level | | | |
|---|--|---|--|---|
| | 1: Start | 2: Low | 3: Medium | 4: High |
| | Preliminary Considerations (PC) | | | |
| PC1. Outcomes – Existence: At what levels of the institution do outcomes exist? | Outcomes are written for a specific segment of the institution, including either courses, programs, departments, schools/colleges, or the institution, but not more than one of these. | Outcomes are written for two specific segments of the institution, including either courses, programs, departments, schools/colleges, or the institution, but only two of these. | Outcomes are written for three specific segments of the institution, including either courses, programs, departments, schools/colleges, or the institution, but only three of these. | Outcomes are written for courses, programs, departments, schools/colleges, and the institution. |
| PC2. Outcomes – Sharing: How are outcomes shared with the institution, its students, and the public? | Outcomes are solely contained in administrative documents used to approve or revise the academic course or program. | Outcomes are shared in limited ways within institutional webpages. | Outcomes of various levels are shared on syllabi, reflected in course descriptions, and published on institutional webpages. | Outcomes are published in multiple sources, including annual reports where data are provided for the external public. |
| PC3. Assessment – Data: What data exist? | Few, if any, measures are used to describe student performance. | A minimal number of measures are used to describe student performance. | Adequate, multiple measures are used to describe student performance. | Multiple, strategically selected, measures are used to triangulate data on student performance. |
| PC4. Assessment – Team: Who performs assessment-related work? | Assessment-related work is predominately performed by an individual. | Assessment-related work is predominately performed by a small group. | Assessment-related work is performed by small to medium sized groups. | Assessment-related work is performed by medium to large sized groups. |
| PC5. Assessment – Structure: How are those responsible for assessment-related work defined and organized? | Those responsible for assessment are individuals assigned on an ad hoc, short-term, part-time, or project-oriented basis. | Those responsible for assessment are defined using simple methods. Either one individual is assigned specific responsibilities, or multiple individuals assigned the smaller duties on a smaller scale. | Those responsible for assessment are multiple individuals performing similar tasks in different areas, requiring high levels of coordination to organize the complexity of the interactions. | Those responsible for assessment have undergone significant change based on a structure that can be maintained. |

| Element/Essential Question | Maturity Level | | | |
|---|--|---|--|---|
| | 1: Start | 2: Low | 3: Medium | 4: High |
| P1. Assessment – Values: How is the value of assessment established? | Values are driven by individuals proposing the need for assessment on campus. | Values are implicit, developed informally by a small collection of individuals, and are communicated in a limited fashion within the group. | Plan (P) Values are explicit, developed formally by a selected group of appropriate (i.e., college, school, or university) representatives and shared widely for input and adoption. | Values are mission-driven, identified through reflection on the institution's working culture. Appropriate stakeholders approve the values. |
| P2. Outcomes – Planning: How is planning for assessment done? | The institution rarely, if ever, plans how it conducts assessment on the campus. | The institution plans how it conducts assessment on a limited basis in more informal circles. | The institution plans how it conducts assessment via specific activities performed on a predictable basis. | The institution plans how it conducts assessment through systematic activities and events. |
| P3. Assessment – Levels: At what levels of the institution does assessment most commonly occur? | Assessment is used to meet individual or course level needs. | Assessment is used to meet individual, course, and program level needs. | Assessment is used to meet individual, course, and program or academic area (e.g., department or school) needs. | Assessment is used to meet needs from individual to institutional. |
| P4. Assessment – Operational Decisions: How are decisions and tasks delegated? | Operational assessment decisions are centralized from one office or person. Delegation is minimal or non-existent. | Operational assessment decisions are delegated on a limited basis. Work to perform exceeds the capacity to delegate. | Operational assessment decisions are hierarchical from the executive level. Work to perform is delegated per the organizational chart. | Operational assessment decisions are distributed throughout the organization parallel to the governance structure. Work to perform is delegated to individuals and groups critical to the governance process. |
| P5. Assessment – Leadership Style: What is the assessment leadership style? | Assessment leadership is unaware of or takes minimal action on the existing problems. | Assessment leadership is driven by survival from one problem to the next. | Assessment leadership identifies and takes action on existing problems and plans for potential problems. | Assessment leadership uses systemic approaches to anticipate, plan, and take action on future problems. |
| B1. Assessment – Culture: Who influences the culture of assessment? | Temporary and/or individual groups influence or establish culture of assessment on an as needed basis (ad hoc). | Smaller, more sustained groups influence or establish culture of assessment based on their needs (working groups). | Build (B) The institution as a whole develops internal approaches to influencing or establishing culture of assessment (hierarchy). | The institution establishes culture of assessment with proper influence from external professional groups (professional culture). |

| Element/Essential Question | Maturity Level | | | |
|---|---|--|--|--|
| | 1: Start | 2: Low | 3: Medium | 4: High |
| B2. Outcomes – Alignment: What alignments exist and how are they used? | Minimal alignments exist internally and are developed only on an as needed basis. | Some alignments exist internally and are developed to meet needs of individual programs. | Alignments exist internally at various levels of the institution to meet needs of multiple programs, and some external alignments exist. | Alignments exist throughout the institution for both internal and external purposes. |
| B3. Assessment – Activities: What is the frequency of assessment activities and their connection to institutional goals? | Assessment activities are not planned, are limited in frequency, or lack connection to institutional goals. | Limited portions of the institution plan assessment activities that occur at an ad hoc frequency and refer to institutional goals. | Significant portions of the institution plan assessment activities that occur at a predictable frequency and connect to institutional goals. | The entire institution consistently plans assessment activities that occur at a systematic frequency and align to institutional goals. |
| B4. Assessment (Staff) – Skill Development: How are the knowledge and skills of the team established and developed? | Assessment staff members are trained according to the minimum institutional expectations. Topics rarely include assessment-specific skills or knowledge. | Assessment staff members are trained according the competencies specific to their position. Few topics include assessment-specific skills or knowledge. | Assessment staff members develop knowledge and skill by comparing, contrasting, and learning each other’s responsibilities (i.e., cross-training). Many topics include assessment-specific skills or knowledge. | Assessment staff members develop diverse expertise as well as knowledge and skill by comparing, contrasting, and learning each other’s responsibilities (i.e., cross-training). A majority of topics include assessment-specific skills or knowledge. |
| B5. Recognition for Engaging in Assessment: How are faculty and staff recognized for engagement in assessment activities? | The institution has no formal system for recognizing engagement in assessment activities where there are no specific types of recognition identified. There is no clear motivation to participate in assessment activities. | The institution has recognitions for generic engagement in assessment activities as part of the established performance monitoring process where there are limited types of recognition used sparingly. Motivation to participate in assessment activities is primarily from extrinsic expectations (i.e., accreditation). | The institution balances generic and strategic incentives for recognizing engagement in assessment activities where there is a menu of types of recognition used frequently. Motivation to participate in assessment activities serves extrinsic expectations (i.e., accreditation) while informing intrinsic needs (i.e., course, program improvement). | The institution uses strategic incentives for recognizing specific engagement in assessment activities where there is a menu of types of recognition used systematically. Motivation to participate in assessment activities balances intrinsic needs (i.e., course, program improvement) with extrinsic expectations (i.e., accreditation). |
| I1. Assessment – Data | Assessment data are | Assessment data are | Implement (I) Assessment data are | Assessment data are |

| Element/Essential Question | Maturity Level | | | |
|---|--|---|---|--|
| | 1: Start | 2: Low | 3: Medium | 4: High |
| Flow: How do data get from assessment staff to the various stakeholders? | disseminated on an ad hoc basis. Assessment report content and design are determined for each report. Data reports are sent to the individual internal stakeholder. | disseminated as needed. Assessment report content and design is consistent and is based on specific needs. Data reports are sent to internal stakeholders (e.g., e-mail). | disseminated regularly to improve accessibility to information. Assessment report content and design is based on a menu of available reports, serving various needs. Data reports are available to internal stakeholders through institutional networks (e.g., the Institutional Research website). | disseminated as part of regular operations of the institution. Assessment report content and design contains both a menu of available reports as well as the ability to make custom reports. Data reports are available to internal and external stakeholders through a combination of networks. |
| I2. Assessment – Role in Accountability: How does use of assessment data influence the institution’s internal or external accountability? | Using assessment data to establish accountability is informal with ad hoc control. Little or no attention is paid to use data to drive program and institutional accountability efforts. | Using assessment data to establish internal accountability is through regularly scheduled diagnostic quality checks. Minimal attention is paid to use data to drive program and institutional accountability efforts. | Using assessment data to establish accountability is focused the achievement of predetermined measures. Data are used to drive program and institutional accountability efforts. | Using assessment data to establish accountability is managed via interactive measures both internal and external to the institution. Data are used to show the vision of the organization is aligned with requirements of the external agency. |
| I3. Technology – Tools: How are tools used to manage the assessment system | No tool is used to document course and program changes. Data has no long term storage strategy. | Existing generic tools (e.g., Microsoft Word or SurveyMonkey) are used to document course and program changes. Data are stored as files on an individual computers or servers | Specific assessment management tools are used to document course and program changes. Data are stored in online databases with the ability to save reports. | Other institutional tools (i.e., student information system) are used with the assessment management system to make course and program documentation more efficient. Data are stored in online databases as part of a larger data management system. |
| I4. Transparency - Definition: How is the institutional definition of transparency determined? | Individuals determine definition of transparency which may be limited in scope. | Departments or larger groups adopt a shared definition of transparency based on the recommendations of other organizations or agencies. | The institution adopts a shared definition of transparency based on the recommendations of other organizations or agencies. | The institution adapts a definition of transparency and then revises it based on data collected. |
| I5. Transparency – Process: How is | Faculty governance process does not discuss or | Faculty governance addresses transparency as issues or | Faculty governance process implements plans for addressing | Faculty governance process implements plans, coordinated |

| Element/Essential Question | Maturity Level | | | |
|--|--|--|--|--|
| | 1: Start | 2: Low | 3: Medium | 4: High |
| transparency managed? | minimally discusses transparency. | opportunities arise. | transparency through an individual committee. | systematically through various committees. |
| I6. Data – Collection: How and when are data collected? | Data are collected manually on an ad hoc basis. | Data are collected manually on an externally imposed (i.e., accreditor) or limited basis. | Data are collected—both manually and automatically—on an ongoing basis. | Data are collected—mostly automatically—on an ongoing basis. |
| E1. Evaluation – Vision: How is the vision of assessment evaluated? | The vision of assessment is unidentifiable. Some goals and actions may be unclear and/or unmeasurable. | The vision of assessment is identified using goals and actions. Goals and actions have measureable elements mainly at the course level. | Evaluate (E) The vision of assessment is clearly developed based on measureable goals and actions. Goals and actions have measureable elements mainly at the program level. | There is a common vision for assessment based on measurable goals and actions. Goals and actions have measureable elements mainly at the institutional level. |
| E2. Communication – Values: How well does the institution communicate the value of assessment? | The institution does not solicit feedback on how it communicates the value of assessment with the campus. | The institution solicits feedback on how it communicates the value of assessment on a limited basis in more informal circles. | The institution solicits feedback on how it communicates the value of assessment via specific faculty development events. | The institution solicits feedback on how it communicates the value of assessment is communicated through on-going faculty development and governance events and mediums. |
| E3. Outcomes – Development: To what degree are various actions and resources used to evaluate the process for developing outcomes? | There have been no or minimal actions to recognize the strengths and weaknesses of how outcomes are developed. There are no resources or no commitments to evaluate the development of outcomes. | Actions are planned to identify strengths and weaknesses at the course or program level. Programs minimally use resources or make commitments to evaluate the development of outcomes. | Reoccurring actions are planned to identify strengths and weaknesses of the process to develop outcomes across many areas of the campus. The institution makes adequate commitment of resources to evaluate the development of outcomes. | Sustained actions are planned to identify strengths and weaknesses of the process to develop outcomes across the entire campus. The institution makes effective commitment of resources to evaluate the development of outcomes. |
| E4. Outcomes – Approval: To what degree are various actions and resources used to evaluate the process for approving outcomes? | There have been no or minimal actions to recognize the strengths and weaknesses of how outcomes are approved. There are no resources or no commitments to | Actions are planned to identify strengths and weaknesses at the course or program level. Programs minimally use resources or make commitments to evaluate approved outcomes. | Reoccurring actions are planned to identify strengths and weaknesses of the process to approve outcomes across many areas of the campus. The institution makes adequate commitment of resources to | Sustained actions are planned to identify strengths and weaknesses of the process to approve outcomes across the entire campus. The institution makes effective commitment of resources to evaluate approved outcomes. |

| Element/Essential Question | Maturity Level | | | |
|--|---|---|--|--|
| | 1: Start | 2: Low | 3: Medium | 4: High |
| E5. Outcomes – Revision: To what degree are various actions and resources used to evaluate the process for revising outcomes? | evaluate approved outcomes. There have been no or minimal actions to recognize the strengths and weaknesses of how outcomes are revised. There are no resources or no commitments to revise outcomes. | Actions are planned to identify strengths and weaknesses are the course or program level. Programs minimally use resources or make commitments to revise outcomes. | evaluate approved outcomes. Reoccurring actions are planned to identify strengths and weaknesses of the process to revise outcomes across many areas of the campus. The institution makes adequate commitment of resources to revise outcomes. | Sustained actions are planned to identify strengths and weaknesses of the process to revise outcomes across the entire campus. The institution makes effective commitment of resources to revise outcomes. |
| E6. Data Discussions: How are data used to stimulate continuous improvement? | Data are rarely, if ever, used in discussion. Data are rarely, if at all, are used for determining achievement of outcomes. Continuous improvement actions are documented rarely, if ever. | Data are used sporadically in conjunction with discussions and other meetings. Data are used to determine meaningful achievement of some outcomes but not all. Continuous improvement actions are documented sporadically. | Data are used periodically as part of a commitment to discuss program improvement. Data are used to demonstrate meaningful achievement of most outcomes. Continuous improvement actions are used regularly. | Data are used routinely to inform continuous improvement at various levels of the institution. Data are used to demonstrate meaningful achievement of all outcomes. Continuous improvement actions are documented routinely. |
| E7. Evaluation – Transparency: To what degree are various actions and resources used to evaluate transparency of assessment efforts? | There have been no or minimal actions to recognize the strengths and weaknesses of how transparency in assessment is demonstrated. There are no resources or no commitments to evaluate transparency of assessment efforts. | Actions are planned to identify strengths and weaknesses of how transparency in assessment is demonstrated at the course or program level. Programs minimally use resources or make commitments to evaluate transparency of assessment efforts. | Reoccurring actions are planned to identify strengths and weaknesses of how transparency in assessment is demonstrated through most, but not all, of the institution. The institution makes adequate commitment of resources to evaluate transparency of assessment efforts. | Sustained actions are planned to identify strengths and weaknesses of how transparency in assessment is demonstrated across the entire campus. The institution makes effective commitment of resources to evaluate transparency of assessment efforts. |

Note. How to Determine Not Applicable: Not Applicable should only be indicated if the purpose and organization of the institution does not conform to any of the choices for reasons beyond lack of evidence. For example, questions regarding voluntary professional accreditation for those programs that do not require it can be marked as Not Applicable. How to Determine No Evidence: No Evidence should only be indicated if the documents used to validate that the institutional commitments to outcomes assessment have never been developed. For example, if the institution has never created alignments among professional standards and no files can be produced as evidence can be marked as No Evidence. It is recommended that any criteria marked as No Evidence have an action plan developed to begin improvement in the area specified.

Appendix B

Comparison of Maturity Models

| Author(s) (Date); Name and Purpose of Maturity Model | Influence on Assessment System Maturity Matrix (ASMM) |
|--|---|
| Humphrey (1987); Capability maturity model: describes software development process | <ol style="list-style-type: none"> 1. Viewing assessment system maturity as a process translates to observable actions that become (rows) of the ASMM 2. The emphasis on routinized tasks is represented in the actions and language used in the ASMM 3. The five phases became the units of performance in the ASMM, with efficient (optimization) being a goal |
| Kendra and Taplin (2004); Change agent competencies model: describes institutional variables that influence maturation | <ol style="list-style-type: none"> 1. The importance of project management skills and competencies translates to how outcomes are implemented; the presence of clear, repeatable plans makes operations more efficient, therefore 2. This reference defined organizational structures of the institution, and influenced how maturity can be observed at the course, program, department, and institutional levels |
| Hammer (2009); Process and enterprise maturity model: describes compatible roles and assets in outcomes assessment system maturity | <ol style="list-style-type: none"> 1. Defining actions based on roles helps the ASMM reflect the various levels of an institution and influences assessment maturity |
| Shepherd (2009); Assessment maturity model: describes the development, delivery, and presentation of assessment results | <ol style="list-style-type: none"> 1. Each criterion in the ASMM was designed from the less ordered and less structured to the more structured 2. Each criterion in the ASMM was written to have observable, documented indicators of maturity |
| Powers (2011); Web content maturity: describes context, channel, consistency, community, consumer (behavior), and criterion of web content | <ol style="list-style-type: none"> 1. The ASMM was designed to document the processes related to how an institution matures its (e.g., the process) of outcomes assessment, not specifically what is communicated (e.g., the data) |
| Solar, Sabattin, and Parada (2013); Maturity model for information and communications technology in school environment: considers the multiple perspectives of stakeholders as they interact with quality of information and technological foundations | <ol style="list-style-type: none"> 1. This reference helped select language in the ASMM that was more observable (less theoretical) and emphasized sustainable, optimized processes related to quality of information and technological |
| Marchewka (2013); Assurance of learning: describes the core actions taken that move the practices from lower to higher maturity in an accredited business school | <ol style="list-style-type: none"> 1. This model influenced the development of the ASMM by going beyond a simple checklist of core actions, emphasizing language more reflective of continuous improvement |
| Aiman and Davis (2014); accommodates maturity at organizational levels and purposes including individual, department, enterprise, optimize, and innovate | <ol style="list-style-type: none"> 1. The ASMM is designed to document assessment work performed at department level first; the results can be aggregated to demonstrate the overall innovation and optimization within the institution 2. The results of departmental use of the ASMM, aggregated into an institutional profile, can be used to stimulate “act local, results global” impact 3. The ASMM is designed to identify matured processes that may serve as centers of assessment work to develop, test, refine, and disseminate various assessment practices throughout the institution |

The *Journal of Educational Research and Practice* provides a forum for studies and dialogue that allows readers to better develop social change in the field of education and learning. Journal content may focus on educational issues of all ages and in all settings. It also presents peer-reviewed commentaries, book reviews, interviews of prominent individuals, and additional content. The objectives: We publish research and related content that examines current relevant educational issues and processes aimed at presenting readers with knowledge and showing how that knowledge can be used to impact social change in educational or learning environments. Additional content provides an opportunity for scholarly and professional dialogue regarding that content's usefulness in expanding the body of scholarly knowledge and increasing readers' effectiveness as educators. The journal also focuses on facilitating the activities of both researcher-practitioners and practitioner-researchers, providing optimal opportunities for interdisciplinary and collaborative thought through blogging and other communications.

Walden University Publishing: <http://www.publishing.waldenu.edu>
