Recontextualizing Bloom's Taxonomy: Quantitative Measures in Formative Curriculum

Assessments and Program Evaluations

Anthony Clemons

Teachers College, Columbia University

Aaron Smith

University of Central Florida

Authors' Note

Anthony Clemons is a Master of Education student at Teachers College, Columbia University and a Curriculum Development Manager with General Dynamics Information Technology. His research and background focuses on methods of designing, developing, and evaluating synchronous and asynchronous instructional systems within andragogical frameworks.

Aaron Smith is a lecturer in the Department of Mathematics at the University of Central Florida. His research addresses the application of statistical theory within educational contexts, particularly with Stochastic Matrices.

corporate affiliations.

Requests for reprints or additional information should be sent to Anthony Clemons, 301 S.

Brett Dr., Ft Knox, KY 40121 or acc2205@tc.columbia.edu and Dr. Aaron Smith, 4393

Andromeda Loop N., MSB 217, Orlando, FL 32816-1364 or aaron.smith@ucf.edu.

Abstract

This article outlines the epistemology, utility, and methodology for formatively evaluating the cognitive achievement of curriculum using a quantifiable assessment process. Building on previous studies, we use Bloom's taxonomy as the quantitative framework for curriculum assessment and implement a three-part evaluation vehicle based on a modified Delphi technique. During Delphi One learning objectives from individual course lessons are coded based on the taxonomical verbiage. In Delphi Two a six-part quantitative curriculum assessment is conducted to determine whether there is progressive cognitive continuity. Delphi Three then requires a holistic curriculum analysis to ensure content alignment with taxonomical objectives and content revisions and objective recoding based on the results of Delphi Two's measures. A quantitative reassessment must then be conducted as one of the final steps of this Delphi to account for those revisions. We conjecture that this technique's utility can only be realized with the use of articulated course learning outcomes and assessable learning objectives for a course's lessons. We recommend this methodology be implemented as part of a curriculum development process and a follow-up study be conducted to determine effectiveness.

Keywords: curriculum evaluation, course assessment, quantifying learning outcomes, weighting, Bloom's Taxonomy, higher order thinking, lower order thinking, critical thinking, Delphi

Recontextualizing Bloom's Taxonomy: Quantitative Measures in Formative Curriculum

Assessments and Program Evaluations

Curriculum evaluations are common practice amongst educators, administrators, and instructional designers within the academic community and industry. Datnow & Hubbard (2015) describe the scope of curriculum evaluations as being concerned with the quality of student interaction during classroom activities, student achievement, student attendance and behavior, course enrollment patterns, postsecondary success rates, and school climate (as cited in Kennedy, 2011; Bernheardt, 1998; Data Quality Campaign, 2011). These research orientations, while certainly necessary, are summative and secondary to the need of a formative assessment of the curricular structure and content being used. In outcomes-based curriculum, a decisive point of content that is assessed are the learning outcomes within a course. James, McInnes, & Delvin (2002) argue that defined mechanisms for measuring and articulating course outcomes must be expected in higher education. This requires "a reasonable surrogate" that can systematically assimilate relevant assessment data through iterative formative evaluations towards demonstrating how a curricula's content and mode of delivery meets its learning outcomes (Horner et al., 2005, p. 48). Doing so improves the propensity for researchers, such as Datnow & Hubbard, to discover higher rates of success in the summative foci they discuss. Yet, research is limited in defining relevant, formulaic curriculum assessment models for analyzing outcomes-based curriculum in interdisciplinary contexts. The introduction of such a model would allow educators to find and interpret meaningful curricular data so that they may evaluate their instructional programs and inform decision makers.

This article outlines the epistemology, utility, and methodology of a formative curriculum assessment method for measuring and quantifying the expected levels of cognition that should be

achieved within a curriculum. While building on previous studies, we continue to use Bloom's taxonomy as the quantitative framework for assessing the cognitive domains that are being achieved. Differentiating this study is our introduction of a modified Delphi technique that uses the taxonomy as a part of a three-phased assessment process; whereby, a curriculum's content is analyzed based on six metrics. These metrics then inform the decisions of program stakeholders as to necessary program overhauls and relevant curriculum refinement needs.

Background and Prior Literature

Of the corpus of relevant research, nearly every study acknowledges *Bloom's Taxonomy of Educational Objectives for the Cognitive Domain* (Bloom et al., 1956) as being a valid framework for quantitatively measuring the effectiveness of curriculum in meeting its intended learning outcomes (Gribble, Meyer, & Jones, 2003; Horner, Zavodska, & Rushing, 2005; Assaly & Smadi, 2015). Armbruster & Ostertag (1989) describe the taxonomy as a hierarchal scheme for identifying a continuum of six specific cognitive demands in learning, ranging from relatively simple (e.g., knowledge, comprehension, application) to more complex (e.g., analysis, synthesis, and evaluation). Assay & Smadi (2015) similarly classify the cognitive domains into two levels of thinking by defining the relatively simple demands as Lower-Order Thinking Skills (LOTS) and the more complex demands as Higher-Order Thinking Skills (HOTS). Through the lens of these thinking skills, this assessment feedback can be critical input for evaluating whether the learning outcomes are adequately aligned, and at what cognitive levels (Krathwohl, 2002). This also provides curriculum designers, textbook writers, and instructors a framework for structuring curriculum with the appropriate distribution of the higher and lower-order cognitive demands (Surjosuseno & Watts, 1999).

Epistemological Underpinnings for Evaluating the Cognitive Domain

Plato intimates in *The Republic* an epistemological distinction between lower and higher forms of learning. He discounts the lower forms as being instructional vehicles for custom, repetition, and "dispositional molding" (Holland, 1980, p. 18). This contrasts his presumption that higher forms of learning foster true enlightenment; whereby, a rational account of individual assertions is required so as to allow for criticism through questioning, towards exposing and extinguishing "erroneous opinion" (p. 21). In his 1923 convocation address, Dartmouth College's late president Earnest Martin Hopkins codified Plato's epistemology by presupposing the aim of higher education as being the cultivation and development of the mind "to the end that [students] may know truth and conform to it" ("Pres. Hopkins outlines the aim of education', 1923, p. 2). This search for truth and expected intellectual comportment is realized through academic coursework that regularly requires students to translate conscious intelligence into action by "distinguish[ing] truth from error" (p. 1).

Higher-order thinking skills (HOTS) in the cognitive domain. Hopkins' sentiments bear the essence of the Platonic epistemology that now has the contemporary distinction of being called Higher-Order Thinking Skills (HOTS). In the hierarchal framework of Bloom's taxonomy, HOTS are achieved by developing curricular content according to the taxonomy's top three cognitive domains of *Evaluation, Synthesis, and Analysis*. The application of HOTS within the cognitive domains can be described as follows:

- Evaluation: This is the highest cognitive form; whereby, a student appraises, assesses, or critiques assignments and exercises based on specific standards and criteria.

- Synthesis: In this role, students bring together interdisciplinary concepts by identifying abstract relations and formulating individual patterns and structures.

- Analysis: In this role, students divide the component parts of ideas and organize them so

they may be understood (Gribble et al., 2005, pp. 10-11).

Assay & Smadi (2015) propose that by incorporating HOTS into a curriculum "students [will be able] to grasp a deep understanding of what they are learning and be more critical and creative instead of merely recalling information" (p. 100). According to Hopkins, this calls for, "the diversity in points of view and [an] emphasis upon stimulating the student's thought" ("Pres. Hopkins outlines the aim of education", 1923, p. 2). Horner et al. (2005) also explain why this pedagogical mechanism matters in the contemporary context, "…[society] expect[s] that students should be able to think critically about the knowledge they inherit…[G]raduates [must] have the ability to analyze fact, data and information and to synthesize and evaluate the "facts" with which they are presented" (p. 1).

Lower-order thinking skills (LOTS) in the cognitive domain. Much like the tenants of the Platonic epistemology, Hopkins also strikes a contrast between lower and higher forms of learning by suggesting the lower forms "[demand] conformity to the thoughts of others" through an emphasis on instruction versus facilitation ("Pres. Hopkins outlines the aim of education", 1923, p. 2). In the contemporary context, curricula delivering instruction using lower learning forms achieves the lower three cognitive domains of *Knowledge, Comprehension, and Application* within Bloom's taxonomy. These domains foster the use of Lower-Order Thinking Skills (LOTS) and can be described as follows:

- Application: In this role, students use new material and apply it through predefined exercises to situations that require them to know and comprehend a subject first.

- Comprehension: In this role, students grasp material through interpretation and articulating estimates based on their knowledge.

- Knowledge: This is the basic cognitive domain; whereby students must recall previously

learned material, including facts, procedures, and theories (Gribble et al., 2005, p. 10).

Dewey (1916) contextualizes LOTS from HOTS by making a distinction *a priori* between the two forms of learning by suggesting alignment with either *training* or *educative teaching*. This distinction is drawn by Dewey's (1916) supposition that training is disciplinary in nature, involving "repeated responses to recurrent stimuli" for the purpose of adjusting cognitive tendencies so individual knowledge might be thwarted to action (p. 36). His supposition supports the premise that lower forms of learning serve individual base faculties for performance in cognitively-light fields, whereas the nature of higher forms of learning tends to be more reflective in cognitive-dense fields. Dewey (1916) goes on to note that industry is innately recalcitrant to maintaining a status quo and regularly changes its products and methods. In turn, industry's changing nature leaves individuals at the mercy of their training, not the virtue of their higher cognitive aptitude. This implies a need for LOTS to be cultivated and nested within a HOTS framework so as to increase autonomous adaptability and versatility when the time arises. Dewey notes that protracting this vulnerability ensures the vitality of mental immaturity, leaving higher faculties under-developed due to the scope of their learning experiences being one of animalistic training versus humanistic education (Dewey, 1916, p. 16). Ultimately, the perpetuation of this industrial pedagogy also inculcates an inequity in opportunity for those whose lives revolve around the continuity of the skills requiring little more than LOTS.

This critique should not be misconstrued to suggest the irrelevance of LOTS in learning. Zohar (2007) argues that achieving the required information for success in academics and life requires a combination of higher and lower thinking skills, with the emphasis placed on tasks requiring higher cognitive demands. Yet, this combination is contingent on students first having a command of certain elementary facts that are both field-specific and interdisciplinary before

being equipped to engage in higher-order thinking (Armbruster & Ostertag, 1989, p. 2). Dewey (1910) describes this as a line of development of logical capacity, whereby students proceed from the concrete to the abstract. This is not to say curricula is limited to a cognitive progression of LOTS-to-HOTS alone. Lesson material can strive to incorporate learning at multiple cognitive levels during a single class and within a single course, without requiring sequential cognitive progression, while achieving course learning outcomes. It is only by assessing the levels and rates at which these cognitive levels are being achieved that stakeholders can holistically evaluate their curricula and make necessary modifications towards achieving the maximum benefit in cognitive participation.

Utility of Bloom's Taxonomy as a Quantifiable Tool

The purpose of assessing curriculum is to discern whether certain goals, outcomes, and objectives are being achieved within a program according to pre-defined parameters (Hong, 2007). A technique in executing a curriculum assessment is to collect data that defines the level of cognition being attained within the content, which then serves as an empirical input for exacting an assessment. While existing research concedes Bloom's Taxonomy as being a valid framework for quantitatively measuring the consistency and quality of HOTS and LOTS, it can also be an input for categorizing and comparing cognitive skills. The taxonomy defines the academic quality of courses through describing course processes "in terms of the level of academic demands or rigor expected of the students" (Nordvall & Braxton, 1996, p. 486). For the taxonomy to effectively convey the expected rigor in an academic context, there must first be the existence of course learning outcomes that have an expected and measurable level of HOTS and LOTS that will be exercised during the course. Without such outcomes being clearly articulated from the outset, a well-defined assessment of the cognitive achievement within a

course cannot be decisively measured, thereby negating the possibility for an accurate programmatic evaluation to take place. By applying the taxonomy as an assessment framework a context of how to describe, categorize, and compare subject content is fostered through a lens of cognitive achievement within six taxonomical domains (Gribble et al., 2005, p. 4).

Nordvall & Braxton (1996) posit that measurements in coursework quality and rigor are also vehicles for distinguishing the quality of one department or institution over another in terms of "academic quality" (p. 487). They limit the validity of any measure of academic quality to several preliminary variables: (a) prior student understanding of the course content, (b) individual student ability, and (c) the instructional goals. They also prescribe that the curricula should be challenging, "but not so challenging that students lack the psychological support they need to meet course expectations" (Nordvall & Braxton, 1996, pp. 487-488). These variables maximize the value of curriculum assessments, making them imperatives for guiding educators in choosing outcomes that are congruent with institutional and departmental goals. They also encourage informed modifications to course materials, thus increasing the likelihood of students achieving an appropriate cognitive level pursuant to the outcomes being sought.

In higher education, defining instructional and cognitive goals is generally hierarchal, beginning at the institutional level and trickling down to the departmental and course levels. It is then at the course level that the learning goals are transposed in the form of outcomes and through those outcomes singular lesson objectives are defined.

(Insert Figure 1)

These outcomes must be defined according to the cognitive levels that will be achieved by the lesson content and instructional techniques of the class session. Figure 1 shows an expected linear progression of the learning objectives of each lesson within a course over a period of 10

lessons. Concurrently, the course learning outcomes are expected to be progressive and sequential, whereby the lesson objectives can be easily nested in the course outcomes, as the Institutional and Departmental goals serve as the overarching framework for which the outcomes and objectives are created. The lesson outcomes and individual objectives are drafted using technical vocabulary that is taxonomically domain-specific in order to competently assess their validity in meeting their cognitive intent. The list of verbs in Table 1 provides an example of the types of qualitative syntax that is needed to assign a quantitative score according to the cognitive domain with which a verb is associated.

(Insert Table 1)

The domains and associated verbs are neither absolute nor discrete but are predominantly useful in describing the cognitive needs associated with course processes (Gribble et al., 2005, p. 4). They also codify the taxonomy's relevance as an assessment framework insofar the terms provide a criterion for formatively assessing curriculum by measurably defining the cognitive actions within lessons.

Formative Curriculum Assessment Methodology

The nature of formative curriculum assessments is that they are conducted prior to curriculum being implemented as an instructional instrument. Relevant literature pertaining methods of formative curriculum assessment generally recommend gathering a few statistical metrics during the development process. While such metrics are certainly relevant for informing the creation of quality curriculum, their application can be maximized in the context of a curriculum assessment strategy that is nested in a programmatic evaluation. By modifying Rowe and White's (1999) characterization of the classical Delphi method, this need can be met through redefining the method's key features as requiring the following:

- Taxonomical assessments of learning outcomes on the merits of pre-defined criteria and expert judgement, in lieu of complete subjectivity on the part of an assessor,

- Iterative curriculum assessments, via a phased program evaluation process that improves the quality of curriculum by the completion of each iteration,

- Short feedback loops between assessors, developers, and administrative stakeholders to provide all possible opportunities for curriculum refinement revisions and input, and

- Quantitative analysis through statistical aggregation of the relevant, pre-defined metrics that can be interpreted during each iteration for improving curriculum quality.

In this case, Figure 2 shows how the Delphi programmatic evaluation process could appear when the method's features are taken into consideration:

(Insert Figure 2)

There are three primary iterations, or rounds, that are introduced in Figure 2, Delphi's R1-3, and several preliminary steps and one post-assessment step that are introduced that contextually influence the assessment measure's in future Delphi rounds. Hsu & Sanford (2007) propose that three rounds of review is usually sufficient in collecting and reviewing information for reaching a consensus by all stakeholders (cf. Cypert and Gant, 1971; Ludwig, 1994, 1997; and Custer, Scarcella, and Stewart,1999). However, with the addition of more rounds, this technique can be both time-consuming and laborious, as it also relies on a lock-step, sequential method and a tremendous amount of input and engagement from process participants. The following discussion provides an overview as to the intricacies of the modified Delphi technique as it is applied to a curriculum assessment process.

Curriculum Development Method Defined

Once all institutional and departmental learning goals have been aligned and learning

stakeholders are in agreement as to the purpose of those goals, course outcomes can then be set. These course outcomes are determined by the intent of the institutional and departmental goals and are formed to be the underlying conceptual framework of the content within a corpus of curriculum. These outcomes are then subdivided into individual lesson learning objectives that define both how the lesson will be taught and what will be learned by the end of the lesson. Ultimately, this drives what Martin-Kniep & Uhrmacher (1992) define as curriculum, or rather "professionally and commercially developed materials" (p. 261). These totality of these materials can include textbooks and units of lesson plans and other instructional products that provide the conceptual framework of a course's outcomes through tying course lessons together in a credible way. This requires a curriculum development process to be decided upon that is iterative and writing-focused, while also concerned with the instructional context in which the learning will occur. Detailing the types of curriculum development models is outside of the scope of this article. However, there are multiple models that can foster the integration of a formative programmatic evaluation process as discussed here.

Delphi Round 1: Coding Learning Objectives

In the first round of the Delphi process, lesson learning objectives are assessed for validity and coded according to the taxonomical domain with which each objective is aligned. Horner et al. (2005) and Assaly & Smadi (2015) use a single-phased curriculum assessment approach, where lesson learning objectives are analyzed and quantitatively coded according to the taxonomical domain each objective achieves based on the highest cognitive verb being used in that objective. Based on this methodology, the individual domain codes are as follows:

Knowledge: 1

Comprehension: 2

Application:	3
Analysis:	4
Synthesis:	5
Evaluation:	6

Once all of the verbs are coded, the sum of instances a verb appears within each of the six domains is computed and can be ranked along a linear spectrum to determine whether "upper-level cognitive skills" are being emphasized and the frequency each domain is being exercised. (Horner et al., 2005 and Assaly & Smadi, 2015). Using Table 3, synchronizing the taxonomical domain coding for a course's learning objectives can be facilitated.

(Insert Table 3)

Specifically, Table 3 allows for a practitioner to nest the learning objectives with learning outcomes within a course, while coding cognitive learning levels towards accruing a measurable sum for future quantitative analysis.

In its current form, the process of taxonomical verb assignment is entirely subjective and lacks a commonly acceptable list of domain-specific verbs that are easily assessable when applied to learning outcomes. This leaves the assessment criterion open to the use of any list of verbs that are believed to be domain-specific, which can result in verbs being used across multiple domains or being abandoned entirely when they might otherwise be relevant. An example of this occurrence can be taken from the lists of verbs used to assess cognitive achievement by The University of West Florida and Marquette University (see Table 1). Upon comparing the two lists, the verb Recognize is applied to the *Knowledge* domain by both universities. However, in the *Comprehension* column *Recognize* is absent from The University of West Florida's verb list but is present in Marquette University's *Comprehension* column

("Bloom's Taxonomy Action Verbs", n.d. and "Action Words for Bloom's Taxonomy", n.d.). While our research did not uncover the practice of verb assignment to be erroneous when applied across multiple domains, there are other verbs which can better describe the outcome being sought by the practitioner. By using the same verb amongst multiple domains, the assessment process becomes arduous due to the assessor having to interpret the correct level of cognition to code, increasing the propensity for error.

Delphi Round 2: Quantitative Curriculum Assessments

The second Delphi round begins with a quantitative assessment of lesson content across the full spectrum of a desired course's curriculum by using the taxonomical codes as the quantitative framework. In statistical analytics the three most important univariate measures correspond to the first three moments of central tendency, spread, and skewness, in that order. Gribble et al. (2005) follow this pattern by proposing the use of four quantitative metrics:

1. The arithmetic mean between each of the six taxonomical cognitive domains that are achieved by each learning outcome for an entire course,

2. The assignment of weight distributions based on the amount of class time allocated to each learning outcome and those of constituent outcomes,

3. The spread of variance from the learning outcomes, relative to their aggregate averages, and

4. The standard deviation from the mean (Gribble et al., 2005, p. 17).

Each of these metrics serve as decisive analytical touchpoints, but for skewness. Therefore, we propose two additional points of assessment:

5. The median, and

6. Pearson's Second Coefficient for Skewness

Each metric adds insight as to cognitive validity and progressivist nature of the lessons. However, it should be noted that these are not the only assessment measures that can be taken and that any of these recommended measures can be assessed through standard computational software. The measures proposed here are suggested due to their simplistic relevance as assessment computations that add tremendous value to towards informing what, if any curriculum improvements might be necessary. Each measure's purpose and computational methodology is discussed in greater detail below.

Computation 1: Arithmetic mean. The arithmetic mean is an estimate of the sampled population's central tendency. In this case the sampled population would be the sum of the taxonomically-coded learning objectives from a course curriculum. This measure is used because it is an unbiased estimator of the sampled population's mean. Furthermore, the Law of Large Numbers guarantees that the arithmetic mean will converge to the population mean when the observations are independent and identically distributed.

The arithmetic mean is found using (Insert Formula 1), where the sum of the scores for each learning objective, x_j, is divided by the total number of the coded learning objectives within a course, n. An example of this computation in practice would be the following two lesson objectives being coded 2 and 4 respectively:

Objective 1. Students will interpret in their own words the purpose of a thesis statement.

Objective 2. Students will be able to distinguish between the different parts of an essay.

The arithmetic mean between the objectives is 3, indicating that the lesson is, on average, aligned within the *Application* taxonomical domain.

Computation 2: Time-weight distributions. Assigning time-weighted distributions to the learning outcomes is given by (Insert Formula 2), where the numerator variable Ω_j represents the

assigned weight (and is a positive number), and x_j represents the value of the coded taxonomical level and the denominator represents the sum of the assigned weights. Using a time-weighted average allows for central tendency of a course's taxonomically coded learning objectives to be found when some objectives are more common than others; thereby allowing certain objectives with more weight to be more influential in the computation. In contrast, the arithmetic mean is the weighted mean, with all objectives being equally influential within the computation. In both the arithmetic and weighted means, if the formula is decomposed into addends, the coefficients of measurements will be positive and sum to one. It follows that the weights in mean calculations form a probability vector. This vector gives stakeholders a convenient way to verify that time weights match the academic program's goals. The arithmetic mean's weights correspond to uniform probability on n possibilities. The sample variance and sample standard deviation formula do not use weights that sum to one; the n-1 in the denominator of sample variance makes it an unbiased estimator of population variance. If the entire population is measured, this n-1 should be replaced with n; for large sample sizes the distinction between n-1 and n is negligible. Therefore, the relevance of this measure extends from the probability that time distributions amongst learning objectives will not always be constant and will vary according to the needs of educational and administrative stakeholders. In short, this measure is used to ensure that the percentage of time spent in a class session on each learning objective is accurately captured using a weighted percentage. This requires a more decisive quantitative measurement that goes beyond the aggregate average of cognitive codes, and gives greater insight into the quality of curriculum design by assessing time as a variable.

In this context, the time is distributed according to the time exercised by each learning outcome during a class. A percentage weight is then found for each learning outcome relative to

the outcome's time distribution and divided by the sum of the weights being assigned. The outcome coded as a four uses ten minutes of the total class-time (0.20); whereas, the outcome coded as a two uses 40 minutes of the class-time (0.80), with the sum of 0.80 and 0.20. Because the weights measure the weighted distribution of time during a class, the sum of the two weights should always equal one, as 100-percent of the class-time distributed between the outcomes. By applying $\mu_w = 0.20(4) + 0.80(2)$, therefore $\mu_w = 2.4$, falling towards the middle of the *Comprehension* domain. By using time-based weighting here, we find that the measure of central tendency and the taxonomical value of the learning objectives changes from where it was in the arithmetic mean.

Computation 3: Variance. The third computational assessment requires the measure of variance (σ^2) relative to the aggregate weighted averages of the learning outcomes. There are two measures of variance, which include the population variance and the weighted variance. Using,

(Insert Population Variance Formula)

the population variance is a quantitative measure of the spread of a course's learning objectives based on the arithmetic mean. The unit of measurement of the variance is the square of observations within a unit of measurement. In this case the observations of the units of measure would be the coded taxonomical objectives. Additionally, the weighted variance is an estimate of the spread of the taxonomical objectives when some observations are more important than others. Observations with more weight are more influential in the computation. The variance's unit of measurement is the square of observations' unit of measurement. Using,

(Insert Weighted Variance formula)

and our example, we find that $\sigma^2 = 0.80(2 - 2.4)^2 + 0.20(4 - 2.4)^2 = 0.64$. This measure allows educators and administrative stakeholders to have an indicator of the squared differences in

distribution values relative to the weighted mean. In turn, this provides them a quantitative basis for consideration as to whether to narrow or widen the variance in how curriculum and learning objectives foster learning.

Computation 4: Standard Deviation. The fourth computational assessment requires the measure of standard deviation (σ). There are two measures of standard deviation, including the population standard deviation and the weighted standard deviation. Using,

(Insert Population Standard Deviation Formula)

the population standard deviation is also a quantitative measure of the spread of a course's learning objectives based on the arithmetic mean. The unit of measurement of the standard deviation is the units of measure observed. In this case the observations of the units of measure would also be the coded taxonomical objectives. Additionally, the weighted standard deviation is an estimate of the population's spread when some observations are more important than others. Observations with more weight are more influential in the computation. Using,

(Insert Weighted Standard Deviation formula)

σ is found by taking the square root of the variance, where $\sigma = 0.80$, resulting in there being a one-to-one relationship between both the variance and standard deviation, pursuant to the stakeholder's preference of which measure to use. However, if either the σ or σ^2 is available the converse can be readily computed.

Computation 5: Median. Measuring for the median of the taxonomically coded learning objectives allows for researchers and practitioners to determine the quality of distribution through knowing what the measure for central tendency is for all of the learning outcomes within a course. This measurement of centrality is also a preferable measure in determining whether the population of course learning objectives skew one way over another. Such a data point would not

otherwise be apparent by using the aggregate average or weighted average. To find the median within a learning outcome data set, the central measure should be found, which would fall between the taxonomic codes of one and six. For example, if the following learning outcomes are coded:

$$1\ 1\ 2\ 3\ 3\ 3\ 3\ 4\ 4\ 4\ 4\ 4\ 5\ 5\ 5\ 6\ 6$$

the median is four due to it being the ninth of 18 learning objectives measured, with an aggregate average of 3.7. However, the median may not always be the best measure due to the propensity for the population of learning objective sets to skew further one way as a result their taxonomical coding. Take the following data set as an example:

$$1\ 2\ 3\ 3\ 5\ 6\ 6\ 6\ 6\ 6\ 6\ 6\ 6\ 6\ 6\ 6$$

If this set were an actual set of learning objectives than the median would be the maximum possible value of six, however, the aggregate average would be 5.1. The nature of this data set makes the aggregate average a more accurate measure. Depending on the distribution of these learning outcomes and the time-weight distributions assigned to each amongst the lessons, the measure could be even lower, making the median an even less useful measure. When data is skewed, a median is considered a more trustworthy measurement of central tendency that means. Medians measure the middle of the data, approximately half of the data will be less than or equal to a median, and approximately half will be greater than or equal. Means can change greatly from the inclusion or exclusion of a few extreme values. Unweighted medians are rank statistics and are resistant to unusually extreme value.

Computation 6: Pearson's Second Coefficient for Skewness.

Skewness measures whether there are observable extreme values following a common direction along either the first and fourth quartiles, incurring an asymmetric form to the central

tendency along that direction. Positive skewed data will have several extreme values greater than the rest of the data. Negative skewed data will have several extreme values less than the rest of the data. Symmetric data will have approximately half of the extreme values on the left, and half on the right when graphically observed. A Measure of Skewness can be assessed, giving researchers an accurate depiction of whether the tendency of the curriculum operates in upper or lower quartiles pursuant to the taxonomical codes assigned to the learning objectives that were analyzed in the first Delphi. One of the easiest ways to spot skewness when analyzing results is to find the difference between the mean and the median. This gives stakeholders a simple arithmetic operation to recognize a lack of symmetry when only sparse statistics are available. Pearson's second coefficient of skewness rescales this difference using,

(Insert Pearson's Second Coefficient Formula Here)

thereby accounting for data spread. If the data is symmetric, the skewness statistic will be close to zero.

Delphi Round 3: Content Analysis, Content Revisions, Learning Objectives Recoded, and a Quantitative Reassessment

Once all of the quantitative metrics are compiled the course being assessed must undergo a comprehensive analysis. During this analysis, lessons endure a thorough review that ensures learning objectives and lesson content actually align. Therefore, if a lesson:

- Fails in meeting the cognitive intent of the taxonomical domain for which the lesson's objective(s) is coded,

- Fails to meet the intended cognitive level that the content reflects within the defined learning objective(s), or

- Fails in being sufficiently progressivist in cognitive achievement,

The lesson must be recomposed by exercising one of the following actions:

- Adjusting the lesson content to meet the actual cognitive level of the objective,

- Adjusting the lesson learning objective(s) to the taxonomically sufficient level that meets the actual cognitive level of the content, or

- Removing the objective and content from consideration entirely.

Once all of the course lessons are analyzed and the appropriate revisions successfully instituted, the actions within Delphi's 1 and 2 must be re-engaged. This ensures the most accurate data is accumulated and is based on layered qualitative and quantitative analytics; thereby giving all stakeholders a decisively precise snapshot of the state of a course's curriculum.

Verifcation and Approval

Though not necessarily a step in the actual assessment process, gaining verification of measures and qualitative feedback of the curriculum under review from neutral, competent sources can be extremely useful. Institutional and departmental stakeholders can and will certainly provide feedback as to curricular expectations in terms of outcomes, and will also ultimately approve the validity and utility of the course measures. However, unbiased feedback from a neutral source informs the process in a holistic sense through providing a "sanity check" as to the alignment of the curriculum, the quality and understandability of the content, and through their insistence on asking elementary questions that might otherwise be ignored by those closely involved.

The process of independent verification also requires the measuring researcher to articulate the process and results in terms that the stakeholders will ultimately require as well. However, it is important to note this process has the potential to be: (1) extremely useful, requiring possible re-writes and concurrent re-assessments based on the quality of feedback given by the

independent party being engaged, or (2) detrimental, with content revisions being incorporated that fail to measure up to the stakeholders' needs, (3) a complete waste of time due to the party's lack of knowledge on the subject(s) and the stakeholders' needs. The process of verification must be definitively contemplated and the process well defined, as well as those who might be involved. Yet, by engaging in this process, the likelihood of stakeholder acceptance of the curriculum measures surges and potential oversights are appreciably mitigated.

Conclusion

The orientation of educational program research is predominantly summative, and focused on collective student achievement and other ancillary factors. Although we acknowledge such *a posteriori* evaluation factors as being relevant in measuring the quality of a course, we must emphasize the utility of a formative evaluation model that measures the quality of a corpus of curriculum *prior* to student engagement. Bloom's Taxonomy, combined with the phased Delphi model we propose provides the "reasonable surrogate" Horner et al. (2005) describe as being necessary in demonstrating how effectively a curricula's content and mode of delivery meets its learning outcomes. Generally, we would appraise effectiveness through the assessment results because they convey the cognitive levels that every lesson should achieve and because the quantitative metric serves as an illustrative benchmark in whether a lesson emphasizes HOTS during facilitation. We would also evaluate whether the lessons and overall course are cognitively progressive and sequential in their structure. Admittingly, the determination of how effectiveness is defined remains within the subjective purview of the stakeholders for whom the evaluation pertains.

Of course, the stakeholders involved in defining effectiveness exercise their intent through specifying strategic instructional and cognitive goals which are ultimately refined by subordinate

stakeholders such as individual departments and course facilitators. However, there may also be extrinsic stakeholders that could come to value the relevance of a quantitative curriculum evaluation methodology in other contexts. Referencing Nordvall & Braxton's (1996) assertion, an evaluation model such as this essentially opens the potential for distinguishing the quality of one department or institution over another in terms of "academic quality" (p. 487). In particular, the result of this potentiality coming to fruition would be the availability of consumer data to prospective parents, students, collegiate ranking agencies, and other relevant parties interested in ascertaining the level of HOTS that a course, department, and institution expects to achieve. This model also affords accreditation agencies the foundational data they would need to articulate how an institution generates "critical thinking", so long as the expectation is set for this evaluation model to be incorporated into the institutional research process.

In closing, this article adds to the limited research field of formative curriculum evaluation processes. We believe it is clear that measuring and quantifying the expected levels of cognition achieved within a curriculum should be integral for informing the decisions of program stakeholders as relevant program overhauls and curriculum refinement needs. We acknowledge that there is a void of empirical research for determining the effectiveness of this model. Therefore, we recommend this methodology be implemented as part of a curriculum development process and a follow-up study be conducted using the proposed modified Delphi technique, so long as the assessment is conducted in concert with taxonomically-articulated lesson learning objectives. We conjecture that only then can this technique's utility can be effectively realized.

ICTCM 28th International Conference on Technology in Collegiate Mathematics

References

"Action Words for Bloom's Taxonomy." University of West Florida, Center for University

Teaching, Learning, and Assessment. Web. 5 Jan. 2016.

<http://uwf.edu/offices/cutla/services-for/assessment/>.

Assaly, I., & Smadi, O. (2015). Using Bloom's taxonomy to evaluate the cognitive levels of

master class textbook's questions. *English Language Teaching*, *8*(5), 100-110.

doi:10.5539/elt.v8n5p100

Bernhardt, V. (1998). *Multiple measures*. Invited Monograph No. 4. Oroville, California:

California Associate for Supervision and Curriculum Development.

Bloom, B., Engleheart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of*

Educational outcomes: The Classification of Educational Goals. Handbook I: Cognitive

domain. New York: Longman.

"Bloom's Taxonomy Action Verbs ." Marquette University. Web. 5 Jan. 2016.

<http://marquette.edu/assessment/includes/documents/BloomsTaxonomyActionVerbs.pd

f>.

Custer, R. L., Scarcella, J. A., & Stewart, B. R. (1999). The modified Delphi technique: A

rotational modification. *Journal of Vocational and Technical Education, 15*(2), 1-10.

Retrieved January 12, 2016, from

https://ejournals.lib.vt.edu/index.php/JCTE/article/view/643/688

Cyphert, F. R., & Gant, W. L.. (1971). The Delphi Technique: A Case Study. The Phi Delta

Kappan, 52(5), 272–273. Retrieved January 6, 2016, from

http://www.jstor.org/stable/20372869

Data: The missing piece to improving student achievement. (2011). Retrieved December 25,

2015, from http://dataqualitycampaign.org/wp-content/uploads/files/dqc_ipdf.pdf

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, *117*(4), 1-26. Retrieved December 25, 2015, from www.tcrecord.org ID Number 17484

Dewey, J. (1910). Chapter 10: Concrete and Abstract Thinking. In *How we think*. New York, New York: D. C. Heath and Company

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education.* New York: Macmillan

Gribble, J., Meyer, L., & Jones, A. (2003). *Quantifying and assessing learning outcomes.*Working Paper SeriesPaper No. 112. Centre for Actuarial Studies, University of Melbourne, Australia. Retrieved from http://repository.unimelb.edu.au/10187/665

Holland, R. (1980). Epistemology and Education. *In Against Empiricism: On Education, Epistemology* and Value. Totowa, New Jersey: Barnes & Noble Books

Hong, S. (2007). *Curriculum evaluation: An evaluation of teaching of science in English in a rural school in Sabah* (Unpublished doctoral dissertation)*.* School of Education and Social Development of the University of Malaysia

Horner, R., Zavodska, A., & Rushing, J. (2005). How challenging? Using Bloom's Taxonomy to assess learning outcomes in a degree completion program. *Journal of College Teaching & Learning, 2*(3), 47-51. http://dx.doi.org/10.19030/tlc.v2i3.1785

Hsu, C., & Sandford, B. (2007). The Delphi technique:Making sense of consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1-7. Retrieved January 6, 2016, from pareonline.net/pdf/v12n10.pdf

James, R., McInnis, C., & Devlin, M. (2002). *Options for a national process to articulate and*

monitor academic standards across Australian universities: Submission to the Higher

Education Review 2002, 1-6. Retrieved December 25, 2015, from

http://www.cshe.unimelb.edu.au/assessinglearning/docs/Crossroads_submission.pdf

Kennedy, M. (2011, May 20). Data use by teachers: Productive improvement or panacea?

Retrieved December 27, 2015, from http://files.eric.ed.gov/fulltext/ED537168.pdf

Krathwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory and Practice,
41*(4), 212-219. http://dx.doi.org/10.1207/s15430421tip4104_2

Ludwig, B. (1994). *Internationalizing extension: an exploration of the characteristics evident in
a state university extension system that achieves internationalization.* (Electronic

Thesis or Dissertation). Retrieved January 9, 2016 from https://etd.ohiolink.edu/

Ludwig, B. (1997). Predicting the future: Have you considered using the Delphi methodology?
Journal of Extension, 35(5), 1-4. Retrieved January 10, 2016 from http://www.joe.org/

joe/1997october/tt2.html

Nordrvall, R., & Braxton, J. (1996). An alternative definition of quality of undergraduate college

education: Toward usable knowledge for improvement. *The Journal of Higher Education,
67*(5), 483-498

Pres. Hopkins outlines the aim of education. (1923, October 4). *Yale Daily News*, pp. 1-8.

Retrieved December 30, 2015, from http://digital.library.yale.edu/cdm/compoundobject/

collection/yale-ydn/id/99658/rec/5

Rowe, G. & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis.
International Journal of Forecasting, 15(4), 353-375

Stanny, C. (n.d.). Action Words for Bloom's Taxonomy. Retrieved January 13, 2016, from http://

uwf.edu/media/university-of-west-florida/offices/cutla/documents/Action-Words-

SLOs-2014-Update-5-5-15.pdf

Surjosuseno, T. T. and Watts, V. (1999). Using Bloom's Taxonomy to teach critical reading in English as a foreign language classes. *Queensland Journal of Educational Research, 15*(2), 227-244. http://education.curtin.edu.au/iier/qjer/qjer15/surjosuseno.html

Zohar, A. (2007). Pedagogical Horizons for Learning. Retrieved January 6, 2016, from http://tlc.cet.ac.il/CETHandler.ashx?n=CetEntities.FileViewer&i=7b6c26a9-ffd0-426d-b508-19d6c9d151f1&id=42650

Table 1
Sample List of Taxonomical Verbs

Higher-Order Thinking Skills			Lower-Order Thinking Skills		
Knowledge	**Comprehension**	**Application**	**Analysis**	**Synthesis**	**Evaluation**
Cite	Add	Acquire	Analyze	Abstract	Appraise
Define	Approximate	Adapt	Audit	Animate	Assess
Describe	Articulate	Allocate	Blueprint	Arrange	Compare
Draw	Associate	Alphabetize	Breadboard	Assemble	Conclude
Enumerate	Characterize	Apply	Break down	Budget	Contrast
Identify	Clarify	Ascertain	Characterize	Categorize	Counsel
Index	Classify	Assign	Classify	Code	Criticize
Indicate	Compare	Attain	Compare	Combine	Critique
Label	Compute	Avoid	Confirm	Compile	Defend
List	Convert	Calculate	Correlate	Construct	Discriminate
Match	Defend	Capture	Detect	Cope	Estimate
Meet	Describe	Change	Diagnose	Correspond	Evaluate
Name	Detail	Classify	Diagram	Create	Explain
Outline	Differentiate	Complete	Differentiate	cultivate	Grade
Point	Discuss	Compute	Discriminate	Debug	Hire
Quote	Distinguish	Construct	Dissect	Depict	Interpret
Read	Elaborate	Customize	Distinguish	Design	Judge
Recall	Estimate	Demonstrate	Document	Develop	Justify
Recite	Example	Depreciate	Ensure	Devise	Measure
Recognize	Explain	Derive	Examine	Dictate	Predict
Record	Express	Determine	Explain	Enhance	Prescribe
Repeat	Extend	Diminish	Figure out	Facilitate	Rank
Reproduce	Extrapolate	Discover	File	Format	Rate
Review	Factor	Draw	Group	Formulate	Recommend
Select	Generalize	Employ	Identify	Generalize	Release
State	Give	Examine	Illustrate	Generate	Select
Study	Infer	Exercise	Infer	Handle	Summarize
Tabulate	Interact	Explore	Investigate	Incorporate	Support
Trace	Interpret	Express	Layout	Integrate	Test
Write	Observe	Factor	Minimize	Lecture	Validate
	Paraphrase	Figure	Order	Model	Verify
	Review	Illustrate	Outline	Modify	
	Rewrite	Investigate	Point out	Network	
	Subtract	Manipulate	Select	Organize	
	Summarize	Modify	Separate	Prepare	
	Translate	Predict	Size p	Prescribe	
		Project	Subdivide	Produce	
		Relate	Transform	Rearrange	
		Sequence		Reorganize	
		Simulate		Rewrite	
		Sketch		Specify	

Table 2
Domain-Specific Verbs Used by Two Universities for Assessing Cognitive Achievement

Knowledge Domain		Comprehension Domain	
Marquette University	Univ. of West Florida	Marquette University	Univ. of West Florida
Arrange	Copy	Classify	Ask
Define	Define	Convert	Associate
Describe	Describe	Defend	Cite
Duplicate	Discover	Describe	Classify
Identify	Duplicate	Discuss	Compare
Label	Enumerate	Distinguish	Contrast
List	Examine	Estimate	Convert
Match	Identify	Explain	Demonstrate
Memorize	Label	Express	Describe
Name	List	Extend	Differentiate
Order	Listen	Generalized	Discover
Outline	Locate	Give example(s)	Discuss
Recognize	Match	Identify	Distinguish
Relate	Memorize	Indicate	Explain
Recall	Name	Infer	Extend
Repeat	Observe	Locate	Estimate
Reproduce	Omit	Paraphrase	Express
Select	Quote	Predict	Generalize
State	Read	Recognize	Give example(s)
	Recall	Rewrite	Group
	Recite	Review	Identify
	Recognize	Select	Illustrate
	Record	Summarize	Indicate
	Repeat	Translate	Infer
	Reproduce		Interpret
	Retell		Judge
	State		Observe
	Select		Order
	Tabulate		Paraphrase
	Tell		Predict
	Visualize		Relate
			Restate
			Report
			Represent
			Research
			Review
			Rewrite
			Select
			Show
			Summarize
			Trace
			Transform
			Translate

Table 3
Curriculum Assessment Matrix for Taxonomical Cognitive Domain Assignments

Course Title	Course Learning Outcome(s)	Lesson Title	Lesson Objective(s)	Taxonomical Cognitive Domain Assignment						Lesson Score
				Level 1: Knowledge	Level 2: Comprehension	Level 3: Application	Level 4: Analysis	Level 5: Synthesis	Level 6: Evaluation	
									Total	

Formula 1. Arithmetic Mean

$$\bar{x} = \frac{\sum\limits_{j=1}^{n} x_j}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Formula 2. Weighted Arithmetic Mean

$$\bar{x}_{weighted} = \frac{\sum\limits_{j=1}^{n} \omega_j x_j}{\sum\limits_{j=1}^{n} \omega_j} = \frac{\omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_n x_n}{\omega_1 + \omega_2 + \ldots + \omega_n}$$

Formula 3. Variance

$$s^2 = \frac{\sum\limits_{j=1}^{n} (x_j - \bar{x})^2}{n-1} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1}$$

Formula 4. Weighted Variance

$$s^2_{weighted} = \frac{\sum\limits_{j=1}^{n} \omega_j (x_j - \bar{x})^2}{\sum\limits_{j=1}^{n} \omega_j} = \frac{\omega_1 (x_1 - \bar{x})^2 + \omega_2 (x_2 - \bar{x})^2 + \ldots + \omega_n (x_n - \bar{x})^2}{\omega_1 + \omega_2 + \ldots + \omega_n}$$

Formula 4. Standard Deviation

$$s = \sqrt{\frac{\sum\limits_{j=1}^{n} (x_j - \bar{x})^2}{n-1}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1}}$$

Formula 5. Weighted Standard Deviation

$$s_{weighted} = \sqrt{\frac{\sum\limits_{j=1}^{n} \omega_j (x_j - \bar{x})^2}{\sum\limits_{j=1}^{n} \omega_j}} = \sqrt{\frac{\omega_1 (x_1 - \bar{x})^2 + \omega_2 (x_2 - \bar{x})^2 + \ldots + \omega_n (x_n - \bar{x})^2}{\omega_1 + \omega_2 + \ldots + \omega_n}}$$

Formula 6. Pearson's Second Coefficient

$$\frac{3(\bar{x} - \tilde{x})}{s}$$

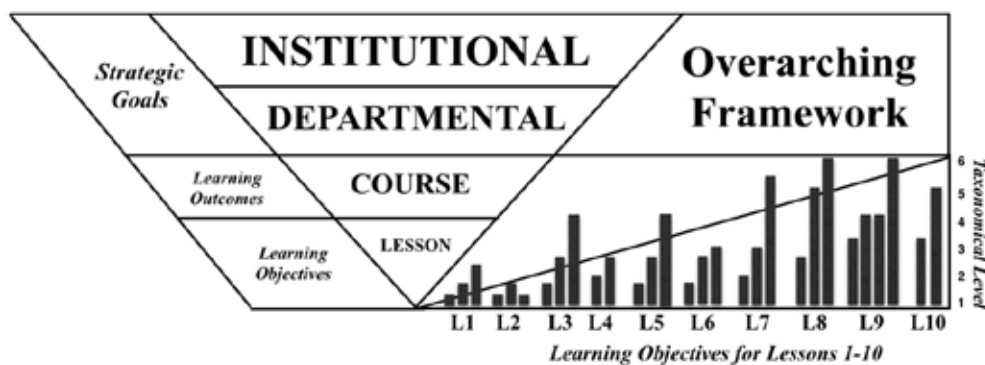Figure 1. Hierarchical Methodology in Defining Cognitive Progression in Learning

Figure 2. Tri-Phased Delphi Evaluation Process for Programmatic Curriculum Assessments