

UC San Diego

UC San Diego Previously Published Works

Title

Probing Internal Assumptions of the Revised Bloom's Taxonomy

Permalink

<https://escholarship.org/uc/item/9087k9x2>

Journal

CBE—Life Sciences Education, 21(4)

ISSN

1931-7913

Authors

Larsen, Tori M
Endo, Bianca H
Yee, Alexander T
et al.

Publication Date

2022-12-01

DOI

10.1187/cbe.20-08-0170

Peer reviewed

Probing Internal Assumptions of the Revised Bloom's Taxonomy

Tori M. Larsen,^{1,2,3,4,†} Bianca H. Endo,^{1,5,†} Alexander T. Yee,^{6,7} Tony Do,^{8,9} and Stanley M. Lo^{1,10,*}

¹Department of Cell and Developmental Biology, School of Biological Sciences, ⁸Preuss School and ¹⁰Program in Mathematics and Science Education, University of California San Diego, La Jolla, CA 92093; ²Program in Biological Sciences, ³Department of Anthropology, and ⁶Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60201

ABSTRACT

Bloom's taxonomy is a classification of learning objectives originally developed for general educational purposes. The taxonomy was revised to expand beyond cognitive processes and to include knowledge types as an orthogonal dimension. As Bloom's taxonomy is a tool widely used in biology education by researchers and instructors, it is important to examine the underlying assumptions embedded within how people may implicitly understand and use the taxonomy. In this paper, we empirically examine two major assumptions: the independence of the knowledge-type and cognitive-process dimensions and the use of action verbs as proxies for different cognitive processes. Contingency analysis on 940 assessment items revealed that the knowledge-type and cognitive-process dimensions are related and not independent. Subsequent correspondence analysis identified two principle axes in how the two dimensions are related, with three clusters of knowledge types and cognitive processes. Using the Shannon evenness index, we did not find a clear relationship between question prompt words (including action verbs) and cognitive processes in the assessment items. Based on these results, we suggest that both dimensions of the revised Bloom's taxonomy should be used and that question prompt words or action verbs alone are not sufficient in classifying the embedded learning objectives within assessment items.

INTRODUCTION

In recent years, reform in undergraduate science education has called for the development of skills such as problem solving and critical thinking to assure the successful transition from education to employment (National Research Council [NRC], 2003, 2007, 2012; National Academies of Sciences, Engineering, and Medicine [NASEM], 2018). As educators, we should work to ensure that classroom assessments are aligned with our educational goals (Wiggins and McTighe, 2005; Astin and Antonio, 2012). Alignment is critical for our understanding of what educational interventions and instructional practices lead to the results we aim to achieve (Handelsman *et al.*, 2007). However, assessments as an aspect of alignment have historically lagged behind other curricular innovations (Orpwood, 2001; Pellegrino, 2013), despite explicit guidance from policy documents.

Researchers have developed a number of classification systems for the purpose of improving curricular alignment. Biggs's structure of the observed learning outcome (SOLO) classifies learning outcomes of student work in terms of complexity (Biggs and Collis, 1982). The Three-Dimensional Learning Assessment Protocol (3D-LAP), developed based on Next Generation Science Standards (NGSS; NRC, 2013), breaks down learning into three dimensions, including scientific and engineering principles, crosscutting concepts, and disciplinary core ideas (Laverty *et al.*, 2016). This study focuses on another commonly used classification system: Bloom's taxonomy (Bloom *et al.*, 1956; Anderson *et al.*, 2001).

Erika Offerdahl, *Monitoring Editor*

Submitted Sep 2, 2020; Revised Jun 1, 2022;
Accepted Jul 13, 2022

CBE Life Sci Educ December 1, 2022 21:ar66
DOI:10.1187/cbe.20-08-0170-CORRECTION

[†]These authors contributed equally to this work. Present addresses: ⁴Freie Universität Berlin, 14195 Berlin, Germany; ⁵Benchling, Inc., San Francisco, CA 94111; ⁷Pressery LLC, Denver, CO 80239; ⁹Department of Communication and Department of Political Science, University of California San Diego, La Jolla, CA 92093.

*Address correspondence to: Stanley M. Lo (smlo@ucsd.edu).

© 2022 T. M. Larsen, B. H. Endo, *et al.* CBE—Life Sciences Education © 2022 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

TABLE 1. Features of the revised Bloom's taxonomy^a

Problems	Revised features
Knowledge as an underlying factor distinct from cognitive processes	Separate knowledge into a dimension orthogonal to the cognitive processes
Direct connections between nonconsecutive categories	Loosen interpretation of the cumulative hierarchical structure in practice
<i>Synthesis</i> and <i>evaluation</i> categories swapped in terms of complexity	Change the order of <i>create</i> and <i>evaluate</i> categories in the cognitive processes
Distinctions among the different categories unclear in practice	Clarify ambiguities by creating extensive subcategories for each category

^aEmpirical evidence demonstrated psychometric inconsistencies in the cumulative hierarchical structure of the original Bloom's taxonomy, which promoted a revision of the taxonomy.

The motivation behind the creation of the original Bloom's taxonomy was to shed light on the behaviors that are important to student learning (Bloom *et al.*, 1956). The taxonomy has been studied extensively and was ultimately revised in 2001 (Anderson *et al.*, 2001). Bloom's taxonomy has been adopted as a research and educational tool by many disciplines, from mathematics to music (Hanna, 2007; Starr *et al.*, 2008; Halawi *et al.*, 2009; Karaali, 2011; Coleman, 2013). In undergraduate biology education, Bloom's taxonomy is widely used in assessing course alignment (Crowe *et al.*, 2008; Jensen *et al.*, 2014), identifying assessment objectives in introductory courses (Momsen *et al.*, 2010, 2013) and mapping out strategies to write effective assessments (Bissell and Lemons, 2006; Lemons and Lemons, 2013; Arneson and Offerdahl, 2018).

Assumptions are embedded within any categorization system, and the meaning and validity of conclusions drawn from a system are situated in such assumptions. Articulating and testing embedded assumptions will help us better understand the tools we are using in research and instruction. For example, Bigg's SOLO levels are assumed to correspond with Piaget's levels of child development in that they both require mastery of earlier stages to master future stages of abstraction (Biggs and Collis, 1989). The 3D-LAP relies on the potential for an assessment item to elicit a particular type of learning in a student, which assumes the features described in their three dimensions are at least somewhat valuable proxies for student cognition (Laverty *et al.*, 2016). For a system as widely used as Bloom's taxonomy, it is important for researchers and instructors to contemplate the underlying assumptions embedded within the structure of the taxonomy, especially as the taxonomy is revised and used in new ways. Here, we articulate and examine key assumptions embedded within how people use Bloom's taxonomy.

Evolution of Bloom's Taxonomy

In the original taxonomy, learning was organized into six cognitive-process categories: *knowledge*, *comprehension*, *application*, *analysis*, *synthesis*, and *evaluation*. (In this paper, we will use italicized words when referring to Bloom's taxonomy categories.) Each category had certain associated cognitive behaviors or action verbs, for example, recalling and recognizing for *knowledge* or defending and judging for *evaluation* (Bloom *et al.*, 1956). However, the cognitive-process categories of the taxonomy were never presented or conceived of as equal subdivisions; rather, they were structured as levels of a cumulative hierarchy organized from simple to complex and from concrete to abstract. This structure assumes a linear progression of the six cognitive-process categories, with links between adjacent levels (e.g., from *knowledge* to *comprehension*) but not between nonadjacent levels (e.g., from *knowledge* to *application*; Stoker

and Kropp, 1971). In other words, performance on *knowledge* questions should directly predict performance on *comprehension* questions but not performance on *application* questions or any other subsequent cognitive-process categories.

Because of the wide reception of the original Bloom's taxonomy, studies examined the underlying assumptions of the cumulative hierarchy structure (Furst, 1981; Kreitzer and Madaus, 1994; Booker, 2007). Empirical evidence based on student performance data indicated that: 1) an underlying factor predicts performance on assessment items in all six cognitive-process categories (Madaus *et al.*, 1973); 2) the *knowledge* category may be part of a different structure (Seddon 1978; Hill and McGaw, 1981); 3) direct connections exist between nonconsecutive categories, such as *comprehension* and *analysis* (Hill and McGaw, 1981); and 4) the *synthesis* and *evaluation* categories are swapped in terms of complexity (Kropp *et al.*, 1966). Furthermore, researchers and instructors alike found it difficult to distinguish between the categories or found the distinctions not helpful (Colder, 1983). These challenges prompted a revision of the taxonomy (Table 1).

The revised taxonomy sought to clarify ambiguities among the different categories in each of the two dimensions by creating extensive subcategories. This expansion of subcategories was done in response to the difficulties of distinguishing the different categories, thus limiting the usability of the original taxonomy (Colder, 1983). For example, instead of simply having an *application* category in the original taxonomy, *apply* in the revised taxonomy is subdivided into *executing* routine procedures in a familiar task versus *choosing* and then implementing a procedure in an unfamiliar task (Anderson *et al.*, 2001). Analogously, *conceptual* knowledge is subdivided into knowledge of classifications and categories, knowledge of principles and generalizations, and knowledge of theories, models, and structures (Anderson *et al.*, 2001). The revised taxonomy suggests that, when categorizing a learning objective or assessment item, one should look to the subcategories for specificity and to help place the objective or assessment in a larger category (Krathwohl, 2002).

Even though the cumulative hierarchy model of the original taxonomy did not have a strong basis in evidence, the revised taxonomy holds steady in its theoretical interpretation that both the knowledge-type and cognitive-process dimensions are organized from simple to complex and from concrete to abstract, perhaps with some potential overlap at either end of each category (Krathwohl, 2002). This conceptualization was motivated by the desire to move from rote learning to meaningful learning, with the assumption that more complex knowledge types and cognitive processes are more meaningful (Anderson *et al.*, 2001). However, the revised taxonomy acknowledges that the cumulative hierarchy is not strict and instead suggests a relaxed

TABLE 2. Parallels among revised Bloom's taxonomy and policy documents^a

Framework	Content	Action	Intersection
Revised Bloom's taxonomy	Knowledge Type	Cognitive process	Learning objective
V&C	Core concepts	Core competencies and disciplinary practices	Biological literacy
NGSS	Disciplinary core ideas, crosscutting concepts	Science and engineering practices	Standards
AP Biology	Big ideas, enduring understanding	Science practices	Performance expectations
MCAT	Foundational concepts, content categories	Scientific inquiry and reasoning skills	Problem solving

^aLearning objectives in the revised Bloom's taxonomy combine the knowledge-type and cognitive-process dimensions and represent the intersection of disciplinary content and actions in context. This structure parallels the frameworks in a number of policy documents, such as *Vision and Change in Undergraduate Biology Education* (V&C), NGSS, AP Biology, and MCAT.

hierarchy (Anderson *et al.* 2001). In practice, mastery of the simple categories may not necessarily be required for mastery of more complex categories (Krathwohl, 2002; Agarwal, 2019). Researchers found empirical evidence for a relaxed hierarchy within the cognitive-process dimension, with overlap occurring among nonadjacent levels (Hill and McGaw, 1981, Agarwal, 2019). Because the assumption of the cumulative hierarchy has been extensively studied within the original and revised taxonomies, we will not be focusing on this assumption in our study. In the following sections, we outline two key features of the revised Bloom's taxonomy and their underlying untested assumptions, which are embedded within how people use the taxonomy in research and practice.

Assumption 1: Independence of Dimensions

The revised taxonomy consists of two dimensions, acknowledging the empirical evidence that knowledge types make an orthogonal dimension to cognitive processes (Seddon 1978; Hill and McGaw, 1981). Any learning objective or assessment item requires some form of knowledge and a cognitive action to perform with that knowledge (Kreitzer and Madaus, 1994; Krathwohl, 2002). The revised taxonomy classifies knowledge into four types: *factual*, *conceptual*, *procedural*, and *metacognitive* (Anderson *et al.*, 2001). Each assessment item should be classified by the intersection of a knowledge type and a cognitive process (Krathwohl, 2002).

In the revised taxonomy, the knowledge-type and cognitive-process dimensions are conceived as independent, with student learning happening at the intersection (Krathwohl, 2002). This structural element is evident in the presentation of every combination of knowledge type and cognitive process as possible. Researchers and instructors alike, especially in biology education, tend to only use the cognitive-process dimension (Allen and Tanner, 2002; Crowe *et al.*, 2008; Momsen *et al.*, 2010, 2013; Freeman *et al.*, 2011; Jensen *et al.*, 2014; Thompson and O'Loughlin, 2015; Semsar and Casagrand, 2017; Lalwani and Agrawal, 2018). The exclusion of the knowledge-type dimension may stem from two potential explanations: 1) Knowledge type does not matter to biology education researchers and instructors for the questions they are asking, and/or 2) people implicitly assume that the two dimensions are so closely associated that only one is needed when using Bloom's taxonomy. Many policy documents involve both knowledge-type and cognitive-process dimensions, suggesting that the first explanation is unlikely to hold true (Table 2). Therefore, it is important to examine the independence of the two dimensions of Bloom's taxonomy, as researchers and educators may implicitly believe that the dimensions are correlated

enough that the cognitive-process dimension alone is sufficient for describing student learning.

Assumption 2: Verbs as Proxies for Cognitive Processes

The revised cognitive-process dimension shifts the categories into their verb forms to emphasize the action focus (Krathwohl, 2002). This dimension includes: *remember*, *understand*, *apply*, *analyze*, *evaluate*, and *create*. To account for some of the psychometric inconsistencies identified in previous studies (Kropp *et al.*, 1966; Seddon 1978; Hill and McGaw, 1981), the new *remember* category takes the place of the original *knowledge* category, and *synthesis* becomes *create* and switches with *evaluate* in terms of complexity.

In the original taxonomy, each category had certain associated cognitive behaviors or action verbs, for example, recalling and recognizing for *knowledge* or defending and judging for *evaluation* (Bloom *et al.*, 1956). In the revision, the cognitive-process categories and subcategories were codified as verbs themselves (Krathwohl, 2002). Perhaps this change reinforces the assumption that verbs could be proxies for their associated cognitive actions. In other words, someone looking to write assessment items in a certain category could simply include the verbs of that category. In this assumption, using the verb "judge" will automatically require students to engage in the cognitive process of *evaluate*. Such associations have continued as a common practice (Stanny, 2016), including attempts to automate classification of assessment items by Bloom's taxonomy using verbs alone (Omar *et al.*, 2012).

Research Questions

The two assumptions, that is, independence of dimensions and verbs as proxies, are both evident in how researchers and instructors use Bloom's taxonomy today and have not been empirically examined. Therefore, our research questions are as follows:

1. Are the knowledge-type and cognitive-process dimensions independent?
2. Can the verbs embedded within assessment items be used as proxies for cognitive processes in Bloom's taxonomy?

METHODS

Data Source

Our data set consisted of a total of 940 assessment items. Of these, 834 were from 12 lower- and upper-division courses across eight biology subdisciplines taught by 16 different instructors in the years 2011–2015 (Figure 1). These courses were offered at a private, not-for-profit, large, primarily

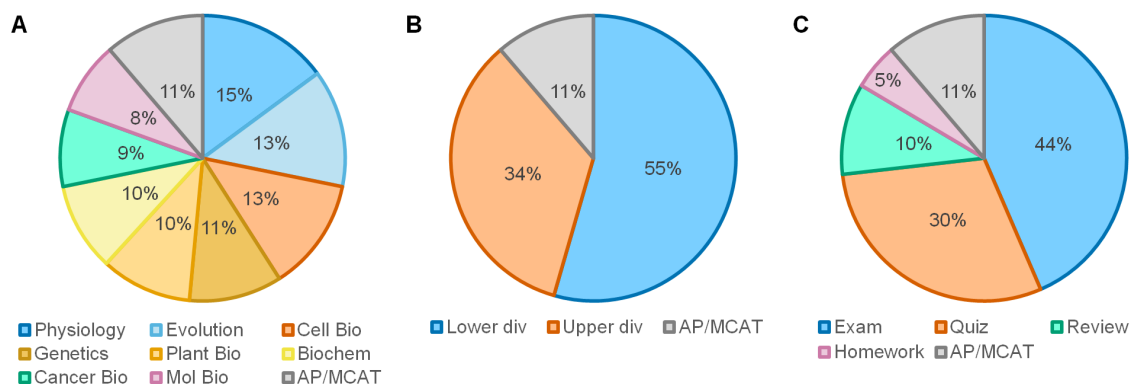


FIGURE 1. Features of our data set. We analyzed assessment items from a variety of biology courses (A); lower- and upper-division courses (B); and assessment sources such as exam, quiz, review, and homework problems (C). The data set also included 51 AP Biology and 55 MCAT questions, which accounted for about 11% of the total sample ($n = 940$). The AP Biology and MCAT questions were combined as one category and included in the pie charts for completeness of the data set.

residential, doctoral university in the midwestern United States, described by the Carnegie Classification of Institutions of Higher Education (McCormick & Zhao, 2005) in the category of highest research activity and with a 4-year, full-time, more selective, and lower transfer-in undergraduate profile. In our preliminary data analysis, we discovered that the assessment items skewed with an overrepresentation of *remember* and *understand* questions. Therefore, another 106 items were added for a total of 940 items. The additional items were included from published sample questions from the Advanced Placement (AP) Biology exam ($n = 51$) and the biological and biochemical sciences section of the Medical College Admission Test (MCAT) exam ($n = 55$). These standardized exams are created through a committee process and had been recently redesigned with the intention to include more assessment items with higher-order cognitive processes (Wood, 2009; Schwartzstein *et al.*, 2013).

Development of the Coding Scheme

The revised Bloom's taxonomy intentionally uses language that is generalizable across contexts and encourages more detailed expansion within individual disciplines (Bloom *et al.*, 1956; Anderson *et al.*, 2001). Therefore, a coding scheme more specific to undergraduate biology assessment items is needed. Our articulation of the taxonomy was a culmination of discussions throughout an iterative coding process, identifying ambiguities to offer an elaboration of the revised taxonomy in the context of biology. To familiarize ourselves with the revised taxonomy, researchers independently generated their own assessment items for each of the six cognitive processes in the revised Bloom's taxonomy and discussed how they would categorize one another's questions. Subsequently, the team categorized the generated items according to the knowledge dimension. All discussions focused on the specific features of these items and their variations that would determine the appropriate categorizations and designations.

After this initial training, the data set was divided in subsets of about 50 assessment items. For each subset, the researchers analyzed the items independently, identifying a knowledge-type category and a cognitive-process category, along with the corresponding subcategories, for each item.

The research team then discussed each item and arrived at consensus for the main categories. Subcategories were not subject to consensus but were used as explanations and points for discussion. In each round of the consensus process, we had extensive discussions on each item to minimize ambiguity in the coding scheme and further delineate features of the various knowledge-type and cognitive-process categories, as well as their subcategories. Our final coding scheme is described in the *Coding Scheme* section.

Context-dependent information can change the way an assessment item is coded. For example, if an instructor passed out a study guide that listed various accepted analyses for a graph, the cognitive process needed to answer a question about that graph would become *remember* instead of *analyze*. In our coding process, only the information given within each assessment item was considered, as we did not have insights into what was discussed in the context of individual courses. The same assumption was made in other studies examining biology assessment items using Bloom's taxonomy (Crowe *et al.*, 2008; Momsen *et al.*, 2013). Groups of related assessment items, such as associated specific diagrams, figures, or models, were kept together for contextual information. These items were coded individually, as each of them may still use different knowledge types and cognitive processes.

Researchers

Four researchers engaged in the iterative coding process. Two were undergraduates (T.M.L. and A.T.Y.) who completed at least half of the introductory biology course sequences for their respective majors at the beginning of the project. We reasoned that undergraduates are especially suited for this type of project, because they are proximal in expertise to students who would encounter these problems on exams or standardized exams. The third researcher (B.H.E.) was a graduate student in biology education research with an undergraduate degree in biological sciences. These three researchers were directly involved in coding assessment items and data analysis. The fourth researcher (S.M.L.) was a biology faculty member with discipline-based education research expertise who engaged in all research discussions.

Reliability

Out of the 940 assessment items in the data set, 17% ($n = 159$) were coded by all three primary researchers, 47% ($n = 442$) by two researchers, and 36% ($n = 339$) by a single researcher. Fleiss' kappa, a generalized form of Cohen's kappa beyond two raters, was used to measure interrater reliability in the subset coded by all three researchers (Fleiss, 1971). Initial interrater reliability for the knowledge-type and cognitive-process categories were $\kappa = 0.68$ and $\kappa = 0.70$, respectively, both falling within the range of substantial agreement (Landis and Koch, 1977). Disagreements within the items with more than one coder were resolved through discussions, and the final consensus was recorded for subsequent data analysis.

Independence of Dimensions

We performed statistical analyses to test whether the knowledge-type and cognitive-process dimensions of the revised Bloom's taxonomy are independent of each other. All statistical analyses were performed in JMP Pro (v. 11.0 to v. 16.0). We used contingency analysis to tabulate the two-dimensional categorical data (i.e., the knowledge-type and cognitive-process categories) from observed frequencies in our data set of biology assessment items and the Fisher's exact test of independence, which is not affected by small numbers, to assess the independence of the two dimensions (Agresti, 1992), as some intersections of knowledge types and cognitive processes have small numbers (Supplemental Table S1). In the Fisher's exact test of independence, the null hypothesis states that the proportions or distributions of one dimension (i.e., cognitive-process categories) are the same across the different values of the other dimension (i.e., knowledge-type categories; McDonald, 2009). When the null hypothesis is rejected, it means that the knowledge-type and cognitive-process categories are related and not independent.

Subsequently, we used correspondence analysis to determine the relationship among the different categories, because our data set contains categorical data. Correspondence analysis is a multivariate method that decomposes the contingency table statistics into orthogonal factors and displays the set of categorical data in a descriptive graphical form, analogous to principal component analysis for continuous data (Jolliffe and Ringrose, 2006). Multidimensional principle axes are calculated to capture as much variation in the data as possible; ultimately, the correspondence analysis results in Cartesian coordinates that denote relationships among categories, where closer categories are considered more related to one another (Greenacre, 2010). Points on the two major dimensions identified in correspondence analyses were put into groups by cluster analysis with hierarchical clustering (Johnson, 1967).

Verbs as Proxies

We expanded our testing of the assumption that verbs within assessment item prompts are being used as proxies for cognitive processes to include question words in addition action verbs. We use "prompt words" as the inclusive term for both. We tabulated specific words used in assessment item prompts as an additional dimension of data to compare with the cognitive-process and knowledge-type categories coded for each assessment item. Some assessment items did not have discrete prompts, consisting of words such as "be," "to," or "as" to infer the ques-

tion being asked. Other assessment items used formatting to imply the questions that students were expected to answer, such as fill in the blanks. For these reasons, 169 assessment items (18% of $n = 940$) were excluded from the prompt word data, resulting in a final sample of 771 items for this analysis. Prompt words were independently recorded by two researchers (B.H.E. and T.D.) who reached consensus for all assessment items.

Shannon evenness index (J') was used to test the hypothesis that prompt words or action verbs can be used as a proxy for cognitive processes. Commonly used to examine biodiversity, J' measures how evenly different species are distributed within an ecosystem by considering the number of species present and the frequency of individuals within each species (Pielou, 1966). Here, we adapt J' to examine the distribution of prompt words in relation to cognitive processes in assessment items. Incidentally, Shannon indices were originally developed to determine the entropy or uncertainty of words in a string of text (Shannon, 1949), suggesting that it is reasonable to use J' to measure the evenness of the distribution of prompt words.

J' ranges between 0 and 1, with higher values signifying a more even distribution of the use of a specific prompt word across all six cognitive processes, and lower values signifying a stronger association of a prompt word to a given cognitive process. J' is calculated with the following formulas (Pielou, 1966):

$$J' = \frac{H'}{\ln(S)} \text{ and } H' = \sum_{i=1}^S p_i \ln(p_i)$$

where S is the number of categories or cognitive processes observed, p_i the proportion of a specific cognitive process used out of the total frequency of a given prompt word, and i the index for the different cognitive processes.

Coding Scheme

In this section, we describe the details of our final coding scheme. Our coding scheme is not meant to be definitive or universal, as course contexts and prior experiences can affect how instructors and students interpret specific learning objectives. Rather, the following description and explanation reflect our interpretation of the revised Bloom's taxonomy. When appropriate, we also clarify distinctions between different categories in the knowledge-type and cognitive-process dimensions with additional frameworks or examples from the existing literature.

Types of Knowledge

Factual and *conceptual* knowledge can be distinguished based on the context of the question (Anderson *et al.*, 2001). We found in our process of delineating these distinctions within our data set that facts consist of a discrete set of details, elements, or specific terminology, whereas concepts draw upon relationships among different facts. Our decision was further informed by the literature on expert versus novice knowledge. Whereas novices tend to see information as isolated facts, experts notice meaningful patterns that connect information (NRC, 2000). The assessment item in Figure 2A asks students about the definition of the different levels of protein structures. To answer the assessment item in Figure 2B, students need to have knowledge not only about the characteristics of k_{cat}/K_M but also how it can

A Factual (Remember)

Which level of protein structure is correctly defined?

- Primary: interaction between subunits of a protein
- Secondary: hydrogen bond arrangement of polar R-groups
- Tertiary: three-dimensional arrangement of all atoms in a single peptide
- Quaternary: order of amino acid residues in the peptide chain
- None of the above are correct

D Understand (Conceptual)

Cholera toxin modifies the G-protein by transferring ADP-ribose to an arginine residue in the GTPase active site. What is the most likely outcome?

- G_q is permanently inactivated because ADP-ribose mimics the GTP-bound state of the protein
- G_q and $G_{\beta\gamma}$ dissociate because of a conformational change induced by the modification
- G_q becomes constitutively activated as the modification prevents GTP hydrolysis
- G_q is released from adenylate cyclase because the modification renders G_q inactive

F Create (Conceptual)

While we have limited ourselves to the coverage of textbook science in class, analogies relevant to biochemistry can be found everywhere in life. Looking at the picture shown, how is what you see relevant to biochemistry? What could this sculpture serve as analogy for? In your answer, identify the specific topic that you think this sculpture could represent, and use 2-3 specific aspects or details to describe in more detail why you think this analogy holds true. In order for you to earn credit here, your answer has to be serious and needs to deliver some real information.

**B Conceptual (Remember)**

Which of the following statements is not characteristic of k_{cat}/K_m ?

- It corresponds to a second-order rate constant
- It provides an excellent parameter for comparison of the catalytic efficiency of enzymes
- It reflects the property of the enzyme when substrate concentration is at saturation
- The upper limit for k_{cat}/K_m is fixed by the diffusion-controlled limit for reactions, which is $10 \text{ M}^{-1}\text{s}^{-1}$

E Analyze (Conceptual)

You discover a protein that inhibits translation. When you add this inhibitor to a mixture capable of translating human mRNA and centrifuge the mixture to separate polyribosomes and single ribosomes, you obtain the following results. The y-axis shows the % of total ribosomes with mRNA attached. Which interpretation is the most consistent with these results?

- The protein binds to the small ribosomal subunit and increases the rate of initiation of translation
- The protein binds to the sequence in the 5' region of the mRNA and inhibits the rate of initiation of translation
- The protein binds to the large ribosomal subunit and slows down elongation of the polypeptide chain
- The protein binds to sequence in the 3' region of the mRNA and prevents termination of translation

C Procedural (Remember)

Which of the following blotting techniques is used to detect RNA?

- Eastern blot
- Northern blot
- Southern blot
- Western blot



FIGURE 2. Example assessment items. We selected examples to highlight differences among knowledge types: *factual* (A), *conceptual* (B), and *procedural* (C); and cognitive processes: *understand* (D), *analyze* (E), and *create* (F). The dimension not highlighted is also included in parentheses for reference.

be used and why it is significant. The multiple-choice option of k_{cat}/K_m “reflect[ing] the property of the enzyme when substrate concentration is at saturation” suggests that students need to know what happens to enzymes when they are saturated with substrates and how this phenomenon relates to the definition of k_{cat}/K_m .

Procedural knowledge consists of information on how and when to use specific skills, algorithms, techniques, or methods; this type of knowledge can be drawn both in theory or practice and can be divided into three subcategories (Anderson *et al.*, 2001). In biology, knowledge of skills and algorithms can include knowing how to read a graph or calculate results using equations. Knowledge of techniques and methods can range from proper pipetting techniques to the scientific method. In addition, procedural knowledge includes criteria for determining when to use appropriate procedures, for example,

knowing when a Western blot is a more appropriate technique than a Northern blot to test a hypothesis. The assessment item in Figure 2C asks students to know which blotting technique is used to detect RNA.

Metacognitive knowledge refers to awareness about oneself and one’s cognition in general, which include strategic knowledge, conditional knowledge, and self-knowledge (Anderson *et al.*, 2001). Students draw upon metacognitive knowledge when thinking about effective test-taking strategies, being aware of theoretical assumptions and experimental limitations, or knowing one’s strengths and weaknesses. Recent studies have placed an emphasis in the role of metacognition in learning and teaching (Bransford *et al.*, 1999; Dauer *et al.*, 2013; NASEM, 2018). However, we did not identify any assessment items in our data set using metacognitive knowledge. In a study on model-based reasoning, students were first tasked to create

a diagram and write a paragraph to explain a phenomenon; subsequently, students were also asked whether they created the diagram or wrote the paragraph first and why (Dauer *et al.*, 2013). The latter prompt represents a question on metacognitive knowledge.

Cognitive Processes

Remember occurs when relevant information is presented with little to no abstraction and is retrieved from memory consistent with how it was originally presented; two subcategories include: recognize and recall (Anderson *et al.*, 2001). To recognize is to identify previously seen information. Recalling involves remembering information with no options to select from. While both subcategories require retrieving information from memory, recalling may be more cognitively demanding (Anderson *et al.*, 2001).

Understand and *analyze* have many similarities (Bloom *et al.*, 1956; Anderson *et al.*, 2001). Even the expanded subcategories in the revised Bloom's taxonomy are seemingly synonymous between the two cognitive processes. For example, classifying, comparing, and explaining as subcategories of *understand* parallel organizing, differentiating, and attributing as subcategories of *analyze*. Such close equivalences remained a constant point of ambiguity and dispute (Elmas *et al.*, 2020). To articulate a clear distinction between the two cognitive processes, we draw inspirations from Biggs's SOLO. Like the original Bloom's taxonomy, Biggs's SOLO is a hierarchical framework: prestructural, unistructural, multistructural, relational, and extended abstract (Figure 3), and these different levels can be applied to classify students' outcomes based on the complexity of their work (Biggs and Collis, 1982). In Biggs's SOLO, students demonstrate the unistructural learning outcome when they are able to make a single generalization between two ideas (Biggs and Collis, 1982). This best corresponded to our articulation of *understand*, a cognitive process that emphasizes the construction of a single relationship. When an assessment item asks students not only to draw multiple connections (multistructure) but also to piece together these relationships in the context of a larger whole (relational), students are asked to *analyze*. The *analyze* cognitive process, therefore, requires the construction of a more complex mental structure, as well as connections on how its constituent parts function together (Dewey, 1933; Zagzebski, 2001). We acknowledge that Biggs's SOLO is intended for students' demonstrated work, whereas Bloom's taxonomy is designed to characterize learning objectives. We do not claim that these two taxonomies are used to measure the same constructs; rather, we borrow from the conceptual organization in Biggs's SOLO as a framework to inform our thinking of how to distinguish the different cognitive processes in the revised Bloom's taxonomy.

The assessment item in Figure 2D prompts students to explain how the structure of a G-protein affects its function. Students use the *understand* cognitive process to determine a single cause-and-effect relationship: Once the active-site residue is modified, the protein function is altered. In contrast, the assessment item in Figure 2E asks students to *analyze* by piecing together multiple connections and the relationships among these connections. To determine the effect of the inhibitor on translation requires answering and connecting the following questions. How are single and polyribosomes related to the pro-

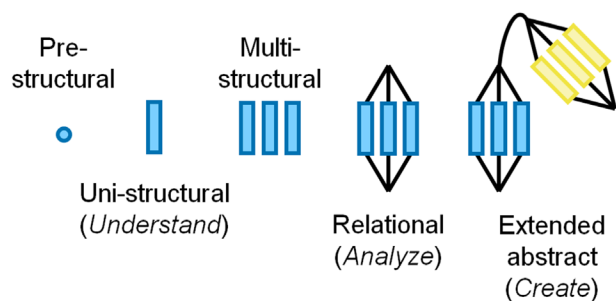


FIGURE 3. Biggs's SOLO as a framework for cognitive processes. Biggs's SOLO describes five levels of complexity in terms of work completed by students. In the prestructural stage, student work tends to be so incomplete that they miss the purpose of the question. There is no Bloom's taxonomy equivalent, as learning objectives should not call for students to miss the point of the question. Unistructural work draws a single connection between two ideas, which parallels the *understand* cognitive process, such as drawing a conclusion based on a direct cause-and-effect connection. Multistructural work demonstrates multiple unistructural connections but does not articulate the relationship among these connections. In our data set, we did not see a Bloom's taxonomy equivalent, as assessment items calling for multistructural outcomes were likely coded as separate *understand* questions. The relational stage, which parallels the *analyze* cognitive process, not only demonstrates multiple connections but also how these connections are related to one another as part of a larger whole. The extended-abstract stage, which parallels the *create* cognitive process, goes beyond the relational stage by connecting the coherent structure to relevant outside information, often bringing in different perspectives and generating novel insights. Figure adapted from Biggs and Collis (1982).

cess of translation? What do they signify in the experimental results? What are the important differences observed when the inhibitor is present or absent? What roles do the small and large subunits, as well as the mRNA, play in the initiation and/or elongation steps of translation? Although both *understand* and *analyze* require drawing relationships, *understand* only requires the establishment of a single relationship, and *analyze* emphasizes how multiple connections work in concert to serve an overall purpose or structure. We note that the presence of experimental data in Figure 2E does not automatically make it an *analyze* question. It is conceivable to rewrite the assessment item in Figure 2D to include experimental data while still asking students to *understand* by determining a single cause-and-effect relationship.

Apply involves using methods or patterns in a given situation, both in theory and in practice, and there are two subcategories: execute and implement (Anderson *et al.*, 2001). In biology, established methods or patterns may include setting up a polymerase chain reaction, using an equation such as Hardy-Weinberg equilibrium, or reading a graph. Executing involves carrying out previously established methods or patterns when encountering a familiar and routine task such as an exercise (Anderson *et al.*, 2001). Implementing entails students working within an existing framework to select a method to solve an unfamiliar problem, and the method is typically more generalized and may have branchpoints of decisions embedded within or multiple position outcomes (Anderson *et al.*, 2001). In principle, the

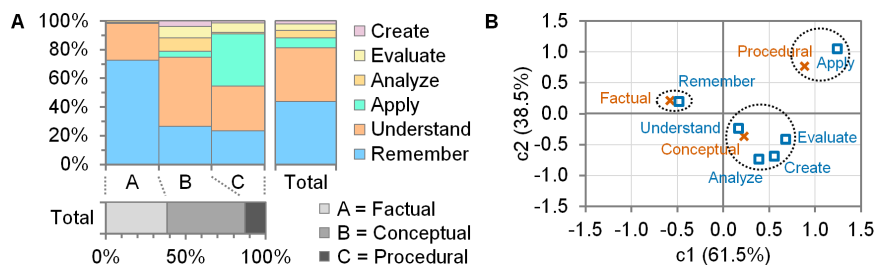


FIGURE 4. Correlation between the knowledge-type and cognitive-process dimensions. (A) Stacked bar graph shows the distribution of assessment items in various combinations of knowledge and cognitive processes. Fisher's exact test reveals that the two dimensions are statistically related ($p < 0.0001$). (B) Correspondence analysis (scatter plot) and hierarchical clustering (dashed ovals) of data reveal three predominant combinations of knowledge types and cognitive processes. The principle axes c1 and c2 accounted for 61.5% and 38.5% of variance in the data, respectively.

framework in question can be based on any type of knowledge in the revised Bloom's taxonomy (Anderson et al., 2001).

The revised Bloom's taxonomy pointed out that the *apply* cognitive process is likely associated with *procedural* knowledge, as this type of knowledge encompasses methods and patterns (Anderson et al., 2001). Both knowledge and cognitive process are used in unison to achieve a specific learning objective. Accessing the knowledge of how or when to do something (i.e., *procedural* knowledge) is different from carrying out a task that uses these established methods or patterns (i.e., *apply*). It is conceivable to have an assessment item that asks students to *remember* specific *procedural* knowledge, for example, to recall different types of microscopy procedures.

Evaluate involves making judgments based on information or evidence, such as the efficiency of a given laboratory method or how consistent a hypothesis is with the supporting evidence. The criteria to determine the certainty or uncertainty of the judgments can be based on a set of metrics specified in the question, in the subcategory of critiquing. Alternatively, the subcategory of checking involves judgments on the internal consistency of the information or evidence, without external metrics being provided. In developing such judgments in the *evaluate* cognitive process, students are asked to make an argument: There is no definitive right or wrong answer, only more or less supported arguments (Anderson et al., 2001).

Create involves organizing multiple components to form a novel, coherent, and functional whole (Anderson et al., 2001). While this echoes some of the ideas in *analyze*, such as taking different elements of a structure into consideration, what sets *create* apart is the requirement that something novel is generated. For example, the assessment item in Figure 2F asks students to generate analogies between biochemistry concepts and features in a sculpture. This example falls under the extended abstract category of Biggs's SOLO (Figure 3), where students are expected to connect multiple relational structures across different frameworks or domains of knowledge. Other forms of *create* use cognitive processes that are typically aligned with activities within scientific investigations. The subcategory of generating involves prompting multiple propositions to explain a given phenomenon, thereby creating new possibilities of knowledge. The subcategories of planning and producing often

go hand in hand, such as designing a series of adaptive experiments to investigate an unknown phenomenon or to arrive at a solution. Like *evaluate*, the *create* cognitive process requires the communication of novel ideas.

RESULTS

The Two Dimensions Are Related and Not Independent

Regarding the knowledge-type dimension, a majority of the assessment items called for *factual* or *conceptual* knowledge (38% and 49%, respectively), with the remaining assessment items calling for *procedural* knowledge (13%; Figure 4A and Table 3). With respect to the cognitive-process dimension, *remember* and *understand* (44% and 37%, respectively) were much more common than other four cognitive processes (Figure 4A and Table 3). Combining both dimensions, *factual* knowledge showed the least variation when associated to the cognitive processes; nearly three-quarters of these assessment items asked students to *remember* (Figure 4A and Table 3). *Conceptual* knowledge was used in combination with the most varied distribution of cognitive processes, accounting for the majority of assessment items calling for the *analyze*, *evaluate*, and *create* cognitive processes, whereas *procedural* knowledge accounted for the majority of assessment items calling for the *apply* cognitive processes (Figure 4A and Table 3).

Fisher's exact test between the knowledge types and cognitive processes revealed a significant relationship between the two dimensions ($p < 0.0001$). When the dimensions are considered together, the combination of *factual* knowledge and *remember* is the most common (28%), followed by the combination of *conceptual* knowledge and *understand* (23%; Figure 4A and Table 3). These observed combinations are more easily visualized through a graphical representation from correspondence analysis followed by hierarchical clustering. The combination of *factual* knowledge and *remember* has the strongest association, signified by the near-direct overlap of both points in the correspondence analysis (Figure 4B). Following hierarchical clustering, our results revealed three predominant combinations of knowledge and cognitive processes: *factual* knowledge with *remember*; *conceptual* knowledge with *understand*, *analyze*, *evaluate*, and *create*; and *procedural* knowledge with *apply* (Figure 4B).

Verbs Are Poor Proxies for Cognitive Processes

For each assessment item, prompt words were recorded to investigate whether such words were associated with specific cognitive processes (Table 3). First, we found that more than half of the assessment items in our data did not use an action verb; instead, 57% of the items used question words such as "which" or "what." Of the five most frequently used prompt words, three were question words ("which," "what," and "how"), and only two were action verbs ("describe" and "explain"). Each of these five prompt words spans at least five out of the six cognitive-process categories and have J' values ranging from 0.58 to 0.85 in our data set (Table 3A). We further

TABLE 3. Shannon evenness indices for most frequently used prompt words^a

A	Remember	Understand	Apply	Analyze	Evaluate	Create	n	J'
Which	85	91	2	8	11	0	197	0.66
What	55	62	25	14	7	0	163	0.85
Describe	30	20	1	4	1	3	59	0.66
How	9	30	5	2	1	0	47	0.66
Explain	6	27	0	1	3	1	38	0.58
B	Remember	Understand	Apply	Analyze	Evaluate	Create	n	J'
Describe	30	20	1	4	1	3	59	0.66
Explain	6	27	0	1	3	1	38	0.58
Name	13	14	0	0	0	0	27	1.00
Draw	0	14	2	0	0	2	18	0.62
Identify	3	5	0	0	0	1	9	0.58

^aPrompt words were recorded for each assessment item, along with the coded type of cognitive process. J' was used to determine the evenness of each prompt word among the six cognitive processes. J' values are reported for the five most frequent prompt words (A) and the five most frequent action verbs (B).

examined the five most frequently used action verbs (“describe,” “explain,” “name,” “draw,” and “identify”), which have J' values ranging from 0.58 to 1.00 in our data set (Table 3B). These high J' values indicate a spread of these prompt words being used across different categories and the lack of association between a given prompt word and a corresponding cognitive process.

DISCUSSION

The findings indicate that, at least in the data set of biology assessment items in our study, the two dimensions of the revised Bloom's taxonomy are related and not independent. We also found that assessment items in biology tend to coincide in certain combinations of knowledge types and cognitive processes. We reason that there may be two potential explanations for these observed combinations. The combinations could be: 1) the result of underlying features of biology as a discipline or 2) the result of the behavioral habits of people, such as instructors, who create assessment items. We further explore these two possibilities in the following paragraphs.

The nature of certain knowledge types and the nature of certain cognitive processes may necessitate that some combinations are statistically more likely to occur in assessment items in biology, similar to how one might expect more *apply* questions in a traditional mathematics course focused on using formulas for calculations. Other combinations, while possible in theory, may not carry practical relevance for the discipline. For example, it may be irrelevant to ever consider making an assessment item that involves creating factual knowledge in biology. Also, while statistically unlikely, some combinations such as *conceptual* knowledge and *apply* were still observed in our data set. Assessment items in these low-frequency combinations may be of particular interest for further examination (Supplemental Figure S1).

It could be that questions involving certain knowledge types and certain cognitive processes are easier or more desirable for instructors to write, making them more likely to occur in assessments. For instance, assessment items that ask students to recall or understand a concept may be easier to write than something that would require a student to create something or evaluate it. Another factor that we were not able to consider in our data was how the format of the question could affect what combinations of knowledge type and cognitive process are more likely.

Both the original and revised Bloom's taxonomy imply that at least cognitive processes may be limited by question format (e.g., multiple choice, short answer, drawing); however, no study has empirically examined this assumption directly (Anderson *et al.*, 2001; Crowe *et al.*, 2008).

Finally, our data demonstrate that prompt words within assessment items are generally not predictive of cognitive processes. For example, while prompt words such as “describe” and “identify” are often considered in relation to the *remember* cognitive process, they can also be associated with the *create* cognitive process in an appropriate context (Figure 2F). In this case, students are asked to identify concepts that they have learned in the course and describe how physical features of the given sculpture are analogies for the course concepts. We conclude that prompt words do not reliably predict cognitive processes; thus, instructors creating assessment items and researchers categorizing them should not simply connect the prompt words in an assessment item with associated verbs and cognitive processes of the revised Bloom's taxonomy. As such, rubrics for researchers and instructors on how to use the revised Bloom's taxonomy should not include a column for associated verbs, despite the form that the cognitive processes of the revised taxonomy takes. Instead, the whole context of the assessment item should be considered.

Limitations

As with earlier studies using Bloom's taxonomy (Crowe *et al.*, 2008; Momsen *et al.*, 2013), we made a similar assumption not to consider course contexts when we coded the assessment items. This means that certain items, if, for instance, they were mentioned word for word in class, may be asking students to use a different set of knowledge types and cognitive processes depending on what was previously presented to students in the course.

Due to the low frequency of certain combinations of knowledge types and cognitive processes in our data set, it is possible that the particular clustering we identified may not be fully generalizable. As such, we do not attempt to make universal claims that are true of all assessment items in biology. However, it is worthwhile to note that across our many instructors and courses, certain combinations remained rare for potential reasons described earlier in the *Discussion*.

While conducting our study, we also revealed another explanation for the dependence of the two dimensions of the

taxonomy that has to do with the procedure of coding. In our methods, we followed convention and coded each assessment item with one knowledge type and one cognitive process corresponding to the presumed learning objective embedded in the assessment item (Bloom *et al.*, 1956; Anderson *et al.*, 2001; Bissell and Lemons, 2006; Lemons and Lemons, 2013; Hanna, 2007; Crowe *et al.*, 2008; Starr *et al.*, 2008; Halawi *et al.*, 2009; Momsen *et al.*, 2010, 2013; Karaali, 2011; Coleman, 2013; Jensen *et al.*, 2014; Arneson and Offerdahl, 2018). However, an assessment item could potentially have multiple embedded or implicit learning objectives. For example, when asking students to analyze experimental data, we may assume that students are already familiar with the technical vocabulary in the assessment item without explicitly labeling such *factual* knowledge as part of the learning objective. Therefore, assigning a single knowledge type and a single cognitive process may not fully capture an instructor's or student's thought process when encountering an assessment item. Instead, we propose that future studies could consider an expanded coding process wherein each assessment item is analyzed for all the embedded types of knowledge and cognitive processes.

Implications

While our data do not suggest that certain combinations of knowledge types and cognitive processes are more or less desirable as learning objectives in undergraduate biology education, we recommend that researchers and instructors who use the revised Bloom's taxonomy should do the following: First, given that there were exceptions to the clustering of certain knowledge types and cognitive processes in the form of rare combinations (Supplemental Figure S1), coding both dimensions would ensure an item is not misclassified. Furthermore, we suggest that instructors who create assessment items should articulate both dimensions for students, as they are both engaging with a type of content (knowledge type) and performing an action (cognitive process). This suggestion is aligned with policy documents (Table 2). Existing Bloom's taxonomy tools in biology (Crowe *et al.*, 2008; Arneson and Offerdahl, 2018) and the 3D-LAP (Laverty *et al.*, 2016) have made important contributions to helping instructors be intentional about the composition of their assessments. Whereas the 3D-LAP aligns well with the NGSS with two dimensions dealing with content and actions, we believe that Bloom's taxonomy tools in biology could be more aligned to policy documents with the inclusion of a content-focused knowledge-type dimension.

Our findings are best understood in light of studies that have shown how the intended and enacted object of learning, as well as what students learn, can be misaligned (Bussey *et al.*, 2013; Dietiker *et al.*, 2018). In this case, learning objectives (intended object of learning) may be misaligned with assessment items (enacted object of learning). Additionally, if an instructor were to assume that a particular verb would indicate a particular cognitive process when it in fact does not, that could result in further misalignment of the intended and enacted objects of learning with what students learn. Prompt words in general should not be used to assume information about the embedded learning objective within an assessment item. Instead, the context of the question and how researchers, instructors, and students approach assessment items should be considered.

Ultimately, a more transparent scaffolding system is needed to support instructors in articulating what they are asking for as students think about how to approach a problem. An expanded coding process like the one we describe may provide such a scaffold. To code all the embedded knowledge types and cognitive processes within an assessment item, both instructors and students would need to articulate the various steps they take within a problem. Then, instructors and students could compare their steps to reveal any potential misalignments.

ACKNOWLEDGMENTS

We are grateful to the faculty participants who contributed assessment items for this analysis. We thank L. Smith and E. Tour for feedback on earlier versions of the article. T.M.L. and A.T.Y. were supported in part by the Undergraduate Research Assistant Program at the Office of Undergraduate Research, Northwestern University. S.M.L. was supported in part by the Faculty Career Development Program at University of California San Diego. This project was initiated with support by an institutional award from the Howard Hughes Medical Institute for undergraduate biology education under award no. 52006934 at Northwestern University. This material is based upon work supported by the National Science Foundation under grant no. DUE-1821724.

REFERENCES

- Agarwal, P. K. (2019). Retrieval practice and Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*(2), 189–209. doi: 10.1037/edu0000282
- Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, *7*(1), 131–153. doi: 10.1214/ss/1177011454
- Allen, D., & Tanner, K. (2002). Approaches to cell biology teaching: Questions about questions. *Cell Biology Education*, *1*(3), 63–67. doi: 10.1187/cbe.02-07-0021
- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Cambridge, MA: Addison Wesley Longman, Inc.
- Arneson, J. B., & Offerdahl, E. G. (2018). Visual literacy in Bloom: Using Bloom's taxonomy to support visual learning skills. *CBE—Life Sciences Education*, *17*(1), ar7. doi: 10.1187/cbe.17-08-0178
- Astin, A. W., & Antonio, A. L. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Lanham, MD: Rowman and Littlefield.
- Biggs, J., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian Journal of Education*, *33*(2), 151–163. <https://doi.org/10.1177/168781408903300205>
- Biggs, J. B., & Collis, K. F. (1982). The psychological structure of creative writing. *Australian Journal of Education*, *26*(1), 59–70. doi: 10.1177/000494418202600104
- Bissell, A. N., & Lemons, P. P. (2006). A new method for assessing critical thinking in the classroom. *BioScience*, *56*(1), 66–72. doi: 10.1641/0006-3568(2006)056[0066:ANMFAC]2.0.CO;2
- Bloom, B. S., Engelhart, M., Furst, M., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals*, Handbook I, *Cognitive domain*. Philadelphia, PA: David McKay Publications.
- Booker, M. J. (2007). A roof without walls: Benjamin Bloom's taxonomy and the misdirection of American education. *Academic Questions*, *20*(4), 347–355. doi: 10.1007/s12129-007-9031-9

- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.
- Bussey, J. T., Orgill, M., & Crippen, J. K. (2013). Variation theory: A theory of learning and a useful theoretical framework for chemical education research. *Chemistry Education Research and Practice*, 14(1), 9–22. doi: 10.1039/C2RP20145C
- Colder, J. R. (1983). In the cells of the "Bloom taxonomy." *Journal of Curriculum Studies*, 15(3), 291–302. doi: 10.1080/0022027830150305
- Coleman, V. M. (2013). *National Music Education Standards and Adherence to Bloom's Revised Taxonomy*. Doctoral dissertation, Minneapolis, MN: Walden University. Retrieved May 31, 2022, from <https://scholarworks.waldenu.edu/dissertations/1048>
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, 7(4), 368–381. doi: 10.1187/cbe.08-05-0024
- Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, 50(6), 639–659. doi: 10.1002/tea.21094
- Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA: D.C. Heath & Co Publishers.
- Dietiker, L., Males, L. M., Amador, J. M., & Earnest, D. (2018). Curricular noticing: A framework to describe teachers' interactions with curriculum materials. *Journal for Research in Mathematics Education*, 49(5), 521–532. doi: 10.5951/jresmetheduc.49.5.0521
- Elmas, R., Rusek, M., Lindell, A., Nieminen, P., Kasapoğlu, K., & Bilek, M. (2020). The intellectual demands of the intended chemistry curriculum in Czechia, Finland, and Turkey: A comparative analysis based on the revised Bloom's taxonomy. *Chemistry Education Research and Practice*, 21(3), 839–851. doi: 10.1039/D0RP00058B
- Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE—Life Sciences Education*, 10(2), 175–186. doi: 10.1187/cbe.10-08-0105
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. <https://doi.org/10.1037/h0031619>
- Furst, E. J. (1981). Bloom's taxonomy of educational objectives for the cognitive domain: Philosophical and educational issues. *Review of Educational Research*, 51(4), 441–453. doi: 10.3102/00346543051004441
- Greenacre, M. (2010). Correspondence analysis of raw data. *Ecology*, 91(4), 958–963. doi: 10.1890/09-0239.1
- Halawi, L. A., McCarthy, R. V., & Pires, S. (2009). An evaluation of e-learning on the basis of Bloom's taxonomy: An exploratory study. *Journal of Education for Business*, 84(6), 374–380. doi: 10.3200/JOEB.84.6.374-380
- Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. New York, NY: Macmillan Publishers Ltd.
- Hanna, W. (2007). The new Bloom's taxonomy: Implications for music education. *Arts Education Policy Review*, 108(4), 7–16. doi: 10.3200/AEPR.108.4.7-16
- Hill, P. W., & McGaw, B. (1981). Testing the simplex assumption underlying Bloom's taxonomy. *American Educational Research Journal*, 18(1), 93–101. doi: 10.3102/00028312018001093
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test ... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26(2), 307–329. doi: 10.1007/s10648-013-9248-9
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241–254. doi: 10.1007/BF02289588
- Jolliffe, I. T., & Ringrose, T. J. (2006). Canonical correspondence analysis. In Kotz, S., Read, C. B., Balakrishnan, N., Vidakovic, B., & Johnson, N. L. (Eds.), *Encyclopedia of statistical sciences*. Hoboken, NJ: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471667196.ess0609>
- Karaali, G. (2011). An evaluative calculus project: Applying Bloom's taxonomy to the calculus classroom. *PRIMUS*, 21(8), 719–731. doi: 10.1080/10511971003663971
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212–218. doi: 10.1207/s15430421tip4104_2
- Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In Anderson, L. W., & Sosniak, L. A. (Eds.), *Bloom's taxonomy: A forty-year retrospective* (pp. 64–81). Chicago, IL: National Society for the Study of Education.
- Kropp, R. P., Stoker, H. W., & Bashaw, W. L. (1966). The validation of the taxonomy of educational objectives. *Journal of Experimental Education*, 34(3), 69–76. doi: 10.1080/00220973.1966.11010940
- Lalwani, A., & Agrawal, S. (2018). Validating revised Bloom's taxonomy using deep knowledge tracing. In *International Conference on Artificial Intelligence in Education* (225–238). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-93843-1_17
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., ... & Cooper, M. M. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol. *PLoS ONE*, 11(9), e0162333. doi: 10.1371/journal.pone.0162333
- Lemons, P. P., & Lemons, J. D. (2013). Questions for assessing higher-order cognitive skills: It's not just Bloom's. *CBE—Life Sciences Education*, 12(1), 47–58. doi: 10.1187/cbe.12-03-0024
- Madaus, G. F., Woods, E. M., & Nuttall, R. L. (1973). A causal model analysis of Bloom's taxonomy. *American Educational Research Journal*, 10(4), 253–262. doi: 10.3102/00028312010004253
- McCormick, A. C., & Zhao, C. M. (2005). Rethinking and reframing the Carnegie classification. *Change: The Magazine of Higher Learning*, 37(5), 51–57. <https://doi.org/10.3200/CHNG.37.5.51-57>
- McDonald, J. H. (2009). Fisher's exact test of independence. *Handbook of biological statistics* (3rd ed., pp. 77–85). Baltimore, MD: Sparky House Publishing.
- Momsen, J., Offerdahl, E., Kryjevskaja, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE—Life Sciences Education*, 12(2), 239–249. doi: 10.1187/cbe.12-08-0130
- Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE—Life Sciences Education*, 9(4), 435–440. doi: 10.1187/cbe.10-01-0001
- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, context, and cultures*. Washington, DC: National Academies Press.
- National Research Council (NRC). (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.
- NRC. (2000). *How People Learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: National Academies Press. <https://doi.org/10.17226/9853>
- NRC. (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.
- NRC. (2012). *Education for life and work: Guide for practitioners*. Washington, DC: National Academies Press.
- NRC. (2013). *Next Generation Science Standards. Next Generation science education: For states, by states*. Washington, DC: National Academies Press.
- Omar, N., Haris, S. S., Hassan, R., Arshad, H., Rahmat, M., Zainal, N. F. A., & Zulkifli, R. (2012). Automated analysis of exam questions according to Bloom's Taxonomy. *Procedia Social and Behavioral Sciences*, 59, 297–303. doi: 10.1016/j.sbspro.2012.09.278
- Orpwood, G. (2001). The role of assessment in science curriculum reform. *Assessment in Education: Principles, Policy and Practice*, 8(2), 135–151. doi: 10.1080/09695940125120
- Pellegrini, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science*, 340(6130), 320–323. doi: 10.1126/science.1232065
- Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13, 131–144. doi: 10.1016/0022-5193(66)90013-0
- Schwartzstein, R. M., Rosenfeld, G. C., Hilborn, R., Oyewole, S. H., & Mitchell, K. (2013). Redesigning the MCAT exam: Balancing multiple perspectives. *Academic Medicine*, 88(5), 560–567. doi: 10.1097/ACM.0b013e31828c4ae0

- Seddon, G. (1978). The properties of Bloom's taxonomy of educational objectives for the cognitive domain. *Review of Educational Research, 48*(2), 303–323. doi: 10.3102/00346543048002303
- Semsar, K., & Casagrand, J. (2017). Bloom's dichotomous key: A new tool for evaluating the cognitive difficulty of assessments. *Advances in Physiology Education, 41*(1), 170–177. doi: 10.1152/advan.00101.2016
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE, 37*(1), 10021. doi: 10.1109/JRPROC.1949.232969
- Stanny, C. J. (2016). Reevaluating Bloom's taxonomy: What measurable verbs can and cannot say about student learning. *Education Sciences, 6*(4), 37. doi: 10.3390/educsci6040037
- Starr, C. W., Manaris, B., & Stalvey, R. H. (2008). Bloom's taxonomy revisited: Specifying assessable learning objectives in computer science. *ACM SIGCSE Bulletin, 40*(1), 261–265. doi: 10.1145/1352135.1352227
- Stoker, H. W., & Kropp, R. P. (1971). An empirical validity study of the assumptions underlying the structure of cognitive processes using Guttman-Lingoes smallest space analysis. *Educational and Psychological Measurement, 31*(2), 469–473. doi: 10.1177/001316447103100215
- Thompson, A. R., & O'Loughlin, V. D. (2015). The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anatomical Sciences Education, 8*(6), 493–501. doi: https://doi.org/10.1002/ase.1507
- Wiggins, G., & McTighe, J. (2005). *Understanding by design, expanded* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- Wood, W. B. (2009). Revising the AP Biology curriculum. *Science, 325*(5948), 1627–1628. doi: 10.1126/science.1180821
- Zagzebski, L. (2001). Recovering understanding. In Steup, M. (Ed.), *Knowledge, truth, and duty: Essays on epistemic justification, responsibility, and virtue*. Oxford, UK: Oxford University Press.