

Using Action Verbs as Learning Outcomes: Applying Bloom's Taxonomy in Measuring Instructional Objectives in Introductory Psychology

Jeffrey S. Nevid & Nate McClelland

Department of Psychology, St. John's University, Queens, New York

Correspondence: Jeffrey S. Nevid, Department of Psychology, St. John's University, 8000 Utopia Parkway, Queens, NY 11439. Tel: 1-718-990-1548. E-mail: nevidj@stjohns.edu

Received: March 11, 2013 Accepted: April 3, 2013 Available online: May 1, 2013

doi:10.11114/jets.v1i2.94

URL: <http://dx.doi.org/10.11114/jets.v1i2.94>

Abstract

We used a set of action verbs based on Bloom's taxonomy to assess learning outcomes in two college-level introductory psychology courses. The action verbs represented an acronym, IDEA, comprising skills relating to identifying, defining or describing, evaluating or explaining, and applying psychological knowledge. Exam performance demonstrated that higher level cognitive skills in Bloom's taxonomy represented by the action verbs *evaluating* and *explaining* were the most difficult for students to acquire and also showed the highest item discrimination index in differentiating between better and poorer students. This study provides a heuristic framework for evaluating areas of relative strength and weakness in acquired skills in college coursework.

Keywords: course assessment, learning goals, Bloom's taxonomy, action verbs

1. Introduction

Academic institutions face increasing pressure to evaluate learning outcomes. All six regional accrediting bodies for higher education in the U.S. require educational institutions to seriously examine educational effectiveness and document student learning (Allen, 2004). A fundamental question underpinning evaluation of learning outcomes is a simple but challenging one: What do we want our students to learn and how can we demonstrate our students are meeting these learning goals?

Various professional groups, including organized psychology, have begun to address the challenge of tying specific learning goals to measurable outcomes (American Psychological Association, 2008; Dunn, McCarthy, Baker, Halonen, & Hill, 2007; Dunn, Mehrotra, & Halonen, 2004). The *APA Guidelines for the Undergraduate Major in Psychology* (2007), which was approved by the American Psychological Association (APA) Council of Representatives in August 2006, specifies ten major learning goals expected of the undergraduate major in psychology. The guidelines offer a set of learning outcomes tied to these particular goals. As the first college level exposure to the discipline of psychology, the introductory psychology course enables students to gain basic skills relating to the acquisition of knowledge of key concepts, theories, and principles consistent with the first learning goal in the APA guidelines, Knowledge Base of Psychology.

The APA guidelines frame learning outcomes in terms of specific skills students are expected to acquire relating to particular learning goals. These learning outcomes are expressed in the form of action verbs, such as analyze, apply, articulate, compare and contrast, define, describe, identify, interpret, demonstrate, distinguish, examine, formulate, evaluate, explain, interpret, locate, recognize, use, and so on. The emphasis on action verbs enables instructors to document the skill sets students are expected to achieve as the result of completing the course. Consider the topic area of biological foundations of behavior, a foundational area in introductory psychology. An example of action verbs applied to acquisition of knowledge about biological bases of behavior might take the form of learning goals such as *identify* parts of the neuron . . . *explain* how an action potential is generated. . . *identify* key neurotransmitters and *describe* their functions . . . and *describe* how the nervous system is organized.

Action verbs are a core feature of the revised version of Bloom's taxonomy (Anderson & Krathwohl, 2001). We selected a set of action verbs to measure instructional objectives representing three levels of cognitive complexity in Bloom's taxonomy based on the listing compiled by Gronlund (1991). We used the verbs *define*, *describe*, and *identify* to measure basic levels of cognitive skills in Bloom's taxonomy (knowledge and comprehension, or remembering and understanding in the revised taxonomy), the verb *apply* to assess an intermediate level of skills development involved in applying knowledge to situations and examples, and the verbs *evaluate* and *explain* to assess higher-order or more complex skills involved in analysis, synthesis, and evaluation (or analyzing and evaluating domains as represented in the revised taxonomy).

2. Purpose of the Present Study

Instructors have long recognized that evaluation of learning outcomes can be used to improve learning and teaching (Halpern, 1988). The measurement of specific types of acquired skills can provide instructors with potentially valuable information regarding areas of relative strength and weakness among students with respect to their acquisition of measured skills.

The purpose of the present study was to perform an item analysis on a set of action verbs used as learning outcomes in introductory psychology courses. The item analysis comprised an examination of item difficulty and item discriminability of examination questions keyed to different types of action verbs, as well as computation of internal consistencies. We used the convenient acronym IDEA to represent four types of action verbs included in the present study: (I) *identifying* key figures in psychology, parts of the nervous system, etc.; (D) *defining* or *describing* key terms and concepts; (E) *evaluating* or *explaining* theoretical constructs and underlying processes and mechanisms; and (A) *applying* concepts to examples.

3. Relevant Scholarship

Efforts to tie learning outcomes to action verbs represent a change from the traditional method of course assessment that focuses on measuring content acquisition to one emphasizing measurement of acquired skills. In the present study, we measured learning outcomes representing a range of acquired skills in introductory psychology courses that broadly related to the hierarchy of skills development represented in Bloom's taxonomy (Bloom, 1956).

Bloom's taxonomy has been widely used across many disciplines to align course objectives and curriculum to level of skills achieved (Dettmer, 2006; Green, 2010; Irish, 1999; Manton, Turner, & English 2004; Su et al., 2005). One frequent observation is that many college courses emphasize rote memorization of factual content or factual minutiae, even though students tend to show poor retention of this type of information (Lord & Baviskar, 2007; Zheng, Lawhorn, Lumley, & Freeman, 2008). Bloom's taxonomy may be helpful to instructors as a scaffolding rubric to help students progress through a hierarchy of skills toward attainment of higher-order cognitive skills, such as applying knowledge and analyzing and evaluating concepts (Athanassiou, McNett, & Harvey, 2003). Instructors may also adapt Bloom's taxonomy to the skills level of their students, as for example, by assigning to the better students more complex tasks involving analysis and synthesis and assigning more basic knowledge and comprehension tasks to the weaker students (Lister & Leaney, 2003). As skills of weaker students develop, they too may be challenged with acquiring more complex cognitive skills.

A major challenge facing instructors seeking to apply a hierarchical skills model such as Bloom's taxonomy is the need to develop reliable and valid means of assessing skills at different levels of cognitive complexity, especially higher-order skills involving more complex cognitive processes (Airasian & Miranda, 2002; Crowe, Dirks, & Wenderoth, 2008). Many different forms of assessment may be used to assess different levels of skills, including multiple choice exams, essay exams, observational techniques, writing assignments, portfolios, and work products (Davis, 2009; Haladyna, 1999; Zepeda, 2007).

Multiple choice tests, which have the advantage of ease of scoring, are widely used to assess student performance in college courses, especially in large introductory or survey courses. Applying Bloom's taxonomy to items on standardized tests such as the AP Biology, biology Graduate Record Exam (GRE), biology section of the Medical College Admission Test (MCAT), as well as course exams in undergraduate biology and basic sciences in medical school, Zheng and colleagues (Zheng et al., 2008) showed that when only multiple-choice questions were considered, the MCAT and GRE contained a higher proportion of higher-order analytical questions than examinations in biology undergraduate courses, AP biology courses, and basic science courses in medical school. Moreover, no significant differences were found in the proportion of higher-level test items between tests that included only multiple choice items (GRE and MCAT) and those that included short-answer or essay questions (AP biology and undergraduate college exams). These findings suggest that multiple choice questions can be used to assess higher-level skills involving analysis and synthesis of basic knowledge and that

efforts are needed to reform undergraduate courses in the sciences to provide greater emphasis on higher-order cognitive skills represented by the upper rungs of the Bloom hierarchy.

4. Methods

The IDEA model of course assessment was incorporated in two introductory psychology courses taught by the same instructor and using the same mainstream textbook at a large, northeastern, metropolitan university. The two classes comprised 144 students in total, 62 males and 82 females. Three non-cumulative, multiple-choice exams were administered, with each exam covering roughly a third of the text. Test items were drawn from the companion test-item file accompanying the textbook and each item was coded for item type (I, D, E, or A in the IDEA model). Test items representing each of the four item types were sampled from each chapter. We measured student performance on each item type by aggregating items across the three exams.

5. Results

5.1 Preliminary Analysis

Classification of item type (I, D, E, or A) was submitted to a blind, interrater reliability study based on a random sample of 50% of test questions, with the two authors serving as independent judges. The results showed an acceptable level of inter-rater agreement (90% concordance). We also computed internal consistencies of the subsets of items comprising each item type by applying the Kuder-Richardson 20 procedure for dichotomous outcomes (correct/incorrect). For all four item types, internal consistencies reached an acceptable level of .70 or higher, with values ranging from .75 to .86. We also found all four item types to be highly interrelated ($p < .01$), with r values ranging from .69 to .83 (see Table 1), which is suggestive of a general cognitive ability or learning factor underpinning performance across item items.

Table 1. Zero-order correlations among IDEA question types

Item Type	1	2	3	4
1. I	1			
2. D	.69*	1		
3. E	.69*	.81*	1	
4. A	.70*	.75*	.83*	1

* $p < .01$

Note. I = Identify; D = Define or Describe; E = Evaluate or Explain, A = Apply

5.2 Item Difficulty

Item difficulty, as represented by the proportion of students answering a test item correctly, was calculated for all test items. Mean difficulty levels for item type are shown in Table 2. A widely-used rule of thumb is to consider items with difficulty levels of less than 20% to be too difficult and those with difficulty levels of greater than 80% to be too easy. As seen in Table 2, the average difficulty levels for the four item types fell within an acceptable moderate range of difficulty (range = .53 to .64).

Table 2. Item difficulty and item discrimination by IDEA question types

Item Type	Item Difficulty	
	<i>M</i>	<i>SD</i>
Identify	.60	.14
Define/Describe	.64	.15
Evaluate/Explain	.53	.15
Apply	.60	.12
Item Type	Item Discrimination Index	
	<i>M</i>	<i>SD</i>
Identify	.29	.17
Define/Describe	.32	.14
Evaluate/Explain	.33	.15
Apply	.27	.17

Note: Difficulty is based on the proportion of students answering items correctly, which is averaged by question

type. The item discrimination index represents the difference between the proportion of students answering an item correctly in the top 27% of the class versus the bottom 27% of the class, averaged by question type.

A repeated measures analysis of variance with the Greenhouse-Geisser correction for nonsphericity revealed a significant overall effect for difficulty level among item types, $F(2.67, 381.82) = 55.1, p < .001, \eta_p^2 = .28$. Consistent with the Bloom hierarchical ordering of acquired skills, follow-up Bonferroni comparisons showed E (Evaluate/Explain) questions to be significantly more difficult than “I” (Identify) questions, $p < .001$, “D” (define or describe) questions, $p < .001$, and “A” (Apply) questions, $p < .001$. Also consistent with the Bloom hierarchy, “A” questions were more difficult than “D” questions, $p < .001$. “I” questions were also more difficult than “D” questions, $p < .001$. Finally, “I” and “A” questions did not differ significantly in difficulty.

5.3 Item Discrimination

Item discrimination is a measure of the ability of test items to discriminate between better and poorer students. We computed an item discrimination index by use of the conventional method of subtracting the proportion of students correctly answering the item in the lowest 27% of the class from the proportion of students correctly answering the item in the highest 27% of the class (Ramsay & Reynolds, 2000). The higher the discrimination index, the better able the items are to discriminate between poorer and better students. Item discrimination indices were aggregated to create an average discrimination index for each question type.

We used three categories for interpreting the index of discrimination based on guidelines suggested by Ebel and Frisbie (1991). Items with a discrimination index of less than .20 were regarded as unacceptably poor discriminators, those between .20 and .29 were considered marginally acceptable discriminators, and those with item discrimination indices of .30 or higher were considered reasonably good discriminators. As shown in Table 2, aggregate item discrimination indices ranged from .27 for Apply questions to .33 for Evaluate/Explain questions. Thus, aggregating test items on the basis of item type (skill assessed) showed that Apply and Identify items were marginally acceptable discriminators, whereas Define/Describe and Evaluate/Explain items met criteria for reasonably good item discrimination.

We also conducted a more discrete item analysis by computing the percentages of individual items within each item set having an item discrimination index of .20 or higher, thus meeting criteria for marginally acceptable or reasonably acceptable discrimination. The percentages of items meeting these criteria were 65% for “I” (Identify) items, 79.4% for “D” (Define or Describe) items, 80.3% for “E” (Evaluate or Explain) items, and 62.3% for “A” (Apply) items. Applying the more stringent criterion of reasonably good discrimination yielded the following percentages of items with each item set, respectively: 52.5% (I), 58.8% (D), 60.6% (E), and 43.5% (A). Although these data point to variability in item discrimination indices both within and across item types, the majority of items for all item types met at least a minimal level of acceptability.

6. Discussion

The present study examined student performance on course examinations in introductory psychology courses with questions coded for type of acquired skills represented by the IDEA acronym: (1) identify, (2) define or describe, (3) evaluate or explain, and (4) apply. Not surprisingly, test performance across item types was highly interrelated, which is reflective of a general cognitive ability factor underlying student performance across different types of test items. Thus, students who do well on one type of test item tend to do well on other item types. However, item types did not appear redundant (correlations were less than .9), which indicates that each type provided distinctive information about student performance.

A hierarchical ordering of skills development consistent with Bloom’s taxonomy was supported by the analysis of item difficulty. As expected, higher order cognitive skills involved in evaluating theoretical concepts or explaining underlying mechanisms or processes proved to be the most difficult of the skills assessed in the present study. Intermediate level skills involved in applying concepts to examples also proved more difficult than basic knowledge acquisition skills involving defining or describing concepts. However, we did not find “apply” questions to be more difficult than “identify” type questions. Nevertheless, the general consistency of our results with the hierarchical ordering of skills development points to the greater challenge instructors face in helping students develop higher order skills needed to evaluate or explain concepts, theories, and underlying processes and mechanisms, as well as to apply concepts to examples rather than to simply define or describe them.

The composite discrimination indices for each item type showed marginally acceptable or reasonably good discrimination between poorer and better students. “Evaluate” and “explain” questions were the most difficult types of items and also attained the highest discrimination index (.33) and the highest percentage of items

(60.6%) reaching a criterion of reasonably good discrimination. The ability to acquire skills relating to evaluation and explanation may be more challenging because it requires more abstract reasoning ability than is the case with other question types. "Evaluate" and "explain" items may also lack available retrieval cues associated with identify, define or describe, or apply types of items. To the extent we would like our students to work with concepts and be able to explain and evaluate the phenomena they study in class presents a continuing challenge to instructors to find better ways of fostering this type of higher learning.

We should also note that multiple choice tests need to be carefully constructed to match the particular skill level the instructor is seeking to assess. In the present study, items were keyed to four types of action verbs representing different skills levels in the Bloom hierarchy. The high internal consistencies of the item sets demonstrated that items assigned to each type were highly interrelated, which indicates that they were measuring a common underlying dimension.

The present study may have heuristic value in providing a method for using action verbs as the basis for assessing learning outcomes in college courses. Instructors can code test items for specific types of skills they would like their students to acquire, using the set of action verbs employed in the present study or drawing upon other inventories of action verbs, such as the listing provided by Gronlund (1995). Analyzing student performance on these item sets may provide useful data indicative of areas of relative skills deficiencies, which instructors can then target by developing teaching strategies designed to strengthen these types of learning outcomes.

The implementation of the course assessment model across two semesters in courses taught by the same instructor helped standardize the instructional materials and examinations across classes. However, a limitation of the present study was that findings may not generalize to classes taught by other instructors or which use other instructional materials.

Instructors recognize the value of assessing development of higher level skills relating to Bloom's taxonomy as students progress through an organized curriculum. Recent evidence from studies in pharmacy education highlight the importance of continuing to address lower level skills while also promoting attainment of higher level skills. In one research example, pharmacy educators evaluated student performance across three exams in an oncology block (Wong, Quist, & Murray, 2006). Students showed no improvement on lower level questions but significant improvement on higher level questions. More recently, pharmacy educators used multiple choice exams to assess student performance at three levels of a modified Bloom's taxonomy (recall, application, and analysis) as students progressed through a course sequence in therapeutics (Tiemeier, Stacy, & Burke, 2011). The results showed poorer student performance through the course sequence on recall items, but improved performance on analysis questions. Taken together, these results underscore the need for educators to continue to focus on attainment of lower level skills as students progress to more advanced topics.

References

- Airasian, P. W., & Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory into Practice, 41*, 249-254. http://dx.doi.org/10.1207/s15430421tip4104_8
- Allen, M. J. (2004). *Assessing academic programs in higher education*. Bolton, MA: Anker.
- American Psychological Association. (2007). *APA guidelines for the undergraduate psychology major*. Washington, DC: Author.
- American Psychological Association. (2008). *Teaching, learning, and assessing in a developmentally coherent curriculum*. Washington, D.C.: American Psychological Association, Board of Educational Affairs. Retrieved from www.apa.org/ed/resources.html
- Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational outcomes: Complete edition*. New York: Longman.
- Athanassiou, N., McNett, J. M., & Harvey, C. (2003). Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. *Journal of Management Education, 27*, 533-555. http://dx.doi.org/10.1177/105256290325251510.1207/s15430421tip4104_8
- Bloom, B.S., (Ed.), Englehart, M.B., Furst, E.J., Hill, W.H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. White Plains, NY: Longman.

- 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, 368-381. <http://dx.doi.org/10.1187/cbe.08-05-0024>
- Davis, B. G. (2009). *Tools for teaching* (2nd ed). San Francisco: Jossey-Bass.
- Dettmer, P. (2006). New Blooms in established fields: Four domains of learning and doing. *Roeper Review*, 28, 70-78. <http://dx.doi.org/10.1080/02783190609554341>
- Dunn, D. S., McCarthy, M., Baker, S., Halonen, J. S., & Hill, G. W. (2007). Quality benchmarks in undergraduate psychology programs. *American Psychologist*, 62, 650-670. <http://dx.doi.org/10.1037%2F0003-066X.62.7.650>
- Dunn, D. S., Mehrotra, C. M., & Halonen, J. S. (Eds.). (2004). *Measuring up: Educational assessment challenges and practices for psychology*. Washington, DC: American Psychological Association.
- Ebel, R. L., & D. A. Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed). Englewood Cliffs, NJ: Prentice Hall.
- Green, K. H. (2010). Matching functions and graphs at multiple levels of Bloom's revised taxonomy. *Primus*, 20, 204-216. <http://dx.doi.org/10.1080%2F10511970802207212>
- Gronlund, N. E. (1995). *How to write and use instructional objectives* (5th ed). New York: Macmillan Publishing Co.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates
- Halpern, D. F. (1988). Assessing student outcomes for psychology majors. *Teaching of Psychology*, 15, 181-186. http://dx.doi.org/10.1207%2Fs15328023top1504_1
- Irish, R. (1999). Engineering thinking: Using Benjamin Bloom and William Perry to design assignments. *Language and Learning across the Discipline*, 3, 83-102.
- Lister, R., & Leaney, J. (2003). Introductory programming, criterion-referencing, and Bloom. *SIGCSE*, 35(1), 143-147. <http://dx.doi.org/10.1145%2F611953.611954>
- Lord, T., & Baviskar, S. (2007). Moving students from information recitation to information understanding: Exploiting Bloom's taxonomy in creating science questions. *Journal of College Science Teaching*, 36, 40-44.
- Manton, E., Turner, C., & English, D. (2004). Testing the level of student knowledge. *Education*, 124, 682-687.
- Ramsay, M. C., & Reynolds, C. R. (2000). Development of a scientific test: A practical guide In M. Hersen & G. Goldstein (Eds.), *Handbook of psychological assessment* (5th ed.) (pp. 21-42). Oxford, UK: Elsevier Science.
- Su, M., Osisek, P. J., & Starnes, B. (2005). Using the revised Bloom's Taxonomy in the clinical laboratory: Thinking skills involved in diagnostic reasoning. *Nurse Educator*, 30, 117-122. <http://dx.doi.org/10.1097%2F00006223-200505000-00014>
- Tiemeier, A. M., Stacy, A. Z., & Burke, J. M. (2011). Using multiple choice questions written at various Bloom's taxonomy levels to evaluate student performance across a therapeutics sequence. *Innovations in Pharmacy*, 2, 1-11. <http://z.umn.edu/INNOVATIONS>
- Wong, S. F., Quist, R., & Murray, W. (2006). Effectiveness of examination blueprints to promote higher cognitive level performance. 107th Annual Meeting of the American Association of Colleges of Pharmacy, San Diego, California, July 9-12. *American Journal of Pharmaceutical Education*, 70(3), 65. <http://dx.doi.org/10.5688/aj700365>
- Zepeda, S. J. (2007). *Instructional supervision: Applying tools and concepts* (2nd ed.). Larchmont, NY: Eye on Education.
- Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's taxonomy debunks the "MCAT myth." *Science*, 319, 414-415. <http://dx.doi.org/10.1126%2Fscience.1147852>

