MULTIPLE-CHOICE: CHOOSING THE BEST OPTIONS FOR MORE EFFECTIVE AND LESS FRUSTRATING LAW SCHOOL TESTING

JANET W. FISHER[*]

I. INTRODUCTION

"I always have trouble with the multiple-choice questions," is a common law student complaint to academic support faculty. Many of the students who seek our assistance are distraught over their failure to perform well on multiple-choice questions because performing poorly on even one test can have serious consequences in the high-stakes testing environment of law school. Doctrinal faculty occasionally express puzzlement to academic support faculty over why some students perform poorly on multiple-choice questions and may refer these students to academic support faculty for help. Academic support faculty find it difficult to give helpful advice to students because these professionals usually are working in the dark as to how doctrinal faculty construct multiple-choice questions. The multiple-choice testing experience could be improved for both faculty and students through adherence to multiple-choice item-writing guidelines and use of appeal or answer justification procedures. These practices would yield more effective and less frustrating testing experiences for both faculty and students.

This article discusses various means by which multiple-choice testing in law school could be improved. Part II provides background on multiple-choice testing. Part III describes problems that arise for students and faculty in connection with multiple-choice testing. Part IV examines possible effects of flawed multiple-choice questions. Part V reviews basic multiple-choice item-writing guidelines. Part VI reviews some general principles of test validity. Finally, Part VII evaluates appeal and answer

justification procedures that could be used to supplement multiple-choice testing.

II. BACKGROUND

Multiple-choice testing, which is widely accepted today as part of the academic landscape, is less than 100 years old.[1] The format was developed as an efficient way to administer intelligence tests to the U.S. Army during World War I.[2] Stanford graduate student Arthur S. Otis is recognized for his contribution to development of the Army intelligence tests.[3] Otis's work, in turn, was influenced by the Kansas Silent Reading Test, developed in 1914 through 1915 by Frederick J. Kelly, which is believed to contain the first known publication of multiple-choice questions.[4]

Use of the new format spread quickly. By 1926 the College Entrance Examination Board had added the multiple-choice SAT (Scholastic Aptitude Tests) to its test offerings.[5] The LSAT (Law School Admission Test) was administered for the first time in l948.[6] The 1948 test took a full day and consisted of ten sections.[7] Most of the sections were in multiple-choice format.[8] While many changes have been made to the LSAT over

---

[1] Franz Samelson, *Was Early Mental Testing: (a) Racist Inspired, (b) Objective Science, (c) A Technology for Democracy, (d) The Origin of the Multiple-Choice Exams, (e) None of the Above? (Mark the RIGHT Answer)*, *in* PSYCHOLOGICAL TESTING AND AMERICAN SOCIETY 1890–1930 113, 115–16 (Michael M. Sokal ed., 1987).

[2] *Id.* at 116.

[3] *Id.* at 116–17.

[4] *Id.* at 118–19.

[5] *Id.* at 122. The SAT as currently administered includes not only multiple-choice questions but also a student-produced essay and other questions that call for student-produced responses. CollegeBoard, SAT Reasoning Test, http://www.collegeboard.com/student/testing/sat/about/SATI.html (last visited Dec. 9, 2008).

[6] LYNDA M. REESE & RUTH ANNE COTTER, A COMPENDIUM OF LSAT AND LSAC-SPONSORED ITEM TYPES, 1948–1994 1 (1994), http://www.lsacnet.org/Research/Compendium-of-LSAT-and-LSAC-Sponsored-Item-Types.pdf.

[7] *Id.* (listing those sections as Verbal Analogies, Sentence Completion, Paragraphs, Word Classification, two sections of Reading Comprehension, Figure Classification, Debates, Contrary and Irrelevant Statements and Reasoning).

[8] *Id.* at 29 app.

the years, the test is still primarily in multiple-choice format.[9] The multiple-choice format also influenced the bar examination process. The National Conference of Bar Examiners created the Multistate Bar Exam (MBE) in the early l970s.[10] The MBE has been a multiple-choice test since its inception.[11]

At the time the MBE was created, law school exams had been dominated by hypothetical essay questions for decades.[12] In response to developments in testing theory, true/false questions were used in law school exams as early as the 1920s, and there are examples of multiple-choice questions on law school exams from the 1950s.[13] With the inception of the MBE, interest in multiple-choice testing grew,[14] and some law faculty began to believe that multiple-choice testing in law school would help prepare students for the MBE.[15] It is likely that law faculty

---

[9] The LSAT as currently administered is a half-day test consisting of five sections of multiple-choice questions, plus a writing sample which is administered at the end of the test. The writing sample is not scored but is sent to each law school to which the examinee has applied. Law School Admission Council, About the LSAT, http://www.lsac.org/LSAT/about-the-lsat.asp (last visited Dec. 15, 2008).

[10] John Eckler, *The Multistate Bar Examination: Its Origins and Objectives*, B. EXAMINER, Feb. 1996, at 14, 16–17.

[11] *Id.* at 17 (noting that the original MBE consisted of 200 multiple-choice questions, covering five subject areas). The MBE as currently administered consists of 200 multiple-choice questions covering six subject areas. National Conference of Bar Examiners, The MultiState Bar Examination (MBE), http://www.ncbex.org/multistate-tests/mbe/ (last visited Dec. 15, 2008).

[12] *See* Steve Sheppard, *An Informal History of How Law Schools Evaluate Students, with a Predictable Emphasis on Law School Final Exams*, 65 UMKC L. REV. 657, 677–79 (1997).

[13] *Id.* at 682–84.

[14] *Id.* at 684.

[15] *See* 1 MICHAEL S. JOSEPHSON, LEARNING AND EVALUATION IN LAW SCHOOL: PRINCIPLES OF TESTING AND GRADING, LEARNING THEORY AND INSTRUCTIONAL OBJECTIVES 39 (1984) [hereinafter VOL. 1 JOSEPHSON] ("Since the vast majority of law school graduates will be required to take the MBE, student's (sic) who have been given no experience with this objective format in their three or four years of law school will be at a distinct disadvantage on the bar exam."); Howard J. Gensler, *Valid Objective Test Construction*, 60 ST. JOHN'S L. REV. 288, 288–89 (1986) ("The objective law school test is valuable both for evaluating the law student and exposing him to multiple[-]choice questions prior to the bar examination.") (citation omitted).

also were attracted to multiple-choice testing because it allowed them to include a greater variety of issues in any one exam with the added bonus of ease of grading.[16]

III. THE PROBLEMS

A variety of problems can arise for students and faculty in connection with multiple-choice testing. Students may incorrectly assess what is required to perform well on multiple-choice questions. Academic support faculty may lack adequate information to advise students how to approach multiple-choice questions. Students who perform poorly on multiple-choice questions are often frustrated and anxious and may want to meet with doctrinal faculty who, in turn, may be puzzled about why particular students performed poorly on the questions.

When students arrive at law school, their most recent experience with multiple-choice testing usually has been the LSAT. Having been successful enough on the LSAT to be admitted to law school, many students believe that the same preparation and test-taking strategies that worked on the LSAT will serve them well on law school multiple-choice exams. Many students describe a multiple-choice test-taking strategy that consists of eliminating two options that "must be wrong" and then reasoning their way to the correct choice between the remaining two options.

Many students do not fully take into account that the LSAT is fundamentally different from law school multiple-choice exams. The LSAT tests general verbal reasoning skills and does not require mastery of any particular body of knowledge.[17] Law school multiple-choice questions, by contrast, are intended to test mastery of legal rules and

---

[16] Gensler, *supra* note 15, at 289, 294.

[17] According to the Law School Admission Council:

> The LSAT is designed to measure skills that are considered essential for success in law school: the reading and comprehension of complex texts with accuracy and insight; the organization and management of information and the ability to draw reasonable inferences from it; the ability to think critically; and the analysis and evaluation of the reasoning and arguments of others.

LSAC.org, About the LSAT, http://www.lsat.org/LSAT/about-the-lsat.asp (last visited Dec. 9, 2008).

concepts. While students seem to understand this difference when it is explained, many have incorrectly assumed that a general familiarity with the legal rules, coupled with their ability to reason their way to a correct answer, will be sufficient to do well on law school multiple-choice exams.

Students must be advised that rule mastery is critical and must be taught ways to achieve that mastery.[18] It is important to emphasize the necessity of working regularly on practice questions in order to engage in the successful application of rules to factual situations that is required to do well on any law school exam—essay or multiple-choice.[19] While this is good advice from academic support faculty, students want to know not only how to prepare for these exams but also how to take them.

Without standard guidelines for the development of law school multiple-choice questions, it is difficult to know what to teach students about these questions. Academic support faculty often find themselves teaching the known styles of particular professors in their law school or working almost completely in the dark if the professors choose not to release any of their questions. Well-intentioned advice from academic support faculty can actually mislead students if the advice is based upon faulty assumptions about how multiple-choice questions on a particular exam will be constructed.[20]

Some academic support advice to students regarding multiple-choice testing assumes faculty are writing multiple-choice items according to guidelines developed by Professor Michael Josephson in his two-volume work on testing in law school developed for the American Association of Law Schools in 1984.[21] Other academic support faculty, likely realizing

---

[18] Rule mastery is built upon sound, thorough and consistent study habits. One critical step toward rule mastery is the development of a rule-based course outline. *See* HERBERT N. RAMY, SUCCEEDING IN LAW SCHOOL 86 (2006) ("When you create a course outline, you are in essence teaching yourself the intricacies of a specific course.").

[19] *See id.* at 112; *see also* MICHAEL HUNTER SCHWARTZ, EXPERT LEARNING FOR LAW STUDENTS 8–9 (2005).

[20] For example, students might be advised to treat each answer choice as an individual true/false question. Such advice overlooks the possibility of a question that calls for relative judgment in which the answer choices will not be absolutely true or false. *See* 2 MICHAEL JOSEPHSON, LEARNING AND EVALUATION IN LAW SCHOOL: TEST CONSTRUCTION, SCORING, GRADING AND RANKING 358 (1984) [hereinafter VOL. 2 JOSEPHSON].

[21] *See* ROGELIO LASSO, TAKING MULTIPLE CHOICE EXAMS 18 n.1, http://www1.law.umkc.edu/Students/BarExamInfo/Taking%20Multiple%20Choice

that they cannot assume that all professors are writing items according to the Josephson guidelines, restrict themselves to more general advice.[22]

Doctrinal faculty are subject matter experts, not experts in testing design and theory,[23] and multiple-choice questions written for law school exams need not be subjected to the same level of scrutiny as questions on standardized exams such as the LSAT or the MBE.[24] Nonetheless, adherence to basic multiple-choice item-writing guidelines would produce more effective exams for students and assist the academic support faculty who work with those students. Furthermore, the use of an appeal or answer justification procedure would likely reduce the amount of time faculty spend in frustrating post-exam conferences with students, improve faculty/student rapport by diminishing anxiety, increase students' sense of autonomy, and provide valuable feedback to faculty regarding student understanding of legal concepts.

IV. POSSIBLE EFFECTS OF FLAWED QUESTIONS

Multiple-choice items[25] that do not adhere to basic multiple-choice item-writing guidelines actually might cause students to fail exams they otherwise would have passed. A study published in 2005 examined the effect of multiple-choice item-writing flaws on medical school examinations.[26] Four faculty-written exams were chosen at random from those given to first and second year medical students enrolled in basic science courses.[27] The test items were reviewed by three judges who independently classified each item.[28] The judges determined that there

---

%20Exams1.pdf (last visited Dec. 15, 2008) (noting that his advice is based upon Michael Josephson's work); *see generally* VOL. 1 JOSEPHSON, *supra* note 15; VOL. 2 JOSEPHSON, *supra* note 20.

[22] *See* SCHWARTZ, *supra* note 19, at 243–46.

[23] VOL. 1 JOSEPHSON, *supra* note 15, at 2 (noting that law teachers rarely have knowledge of principles of test construction or psychometric theory).

[24] *See id.* at 7.

[25] *See* THOMAS M. HALADYNA, DEVELOPING AND VALIDATING MULTIPLE-CHOICE TEST ITEMS 3–4 (3d ed. 2004) (referring to individual test questions as "items").

[26] Steven M. Downing, *The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed Test Items on Achievement Examinations in Medical Education,* 10 ADVANCES HEALTH SCI. EDUC. 133, 134 (2005).

[27] *Id.* at 135.

[28] *Id.*

were 100 flawed items out of the total 219 items on the four exams, and just five item flaws accounted for 90 of the 100 flaws.[29] The author noted, "The item flaws studied were non-subtle, obvious violations of the well[-] established principles of effective multiple-choice item writing."[30]

The flawed items were more difficult for students in three of the four exams studied and had a negative impact on passing rates.[31] The author concluded that, "some students—perhaps as high as 10–15% of students tested—were incorrectly classified as failed when they should have been classified as passed, due solely to flawed item formats and the ineptitude of test item writers."[32] The study recommended further research into the effects of flawed test questions and an increase in efforts to teach medical school faculty multiple-choice item-writing principles.[33]

V. WRITING MULTIPLE-CHOICE ITEMS

It has been noted that writing "good test items is probably the most demanding type of creative writing imaginable."[34] The item writer must understand the content, determine the cognitive behavior to be tested, and write with originality and clarity.[35] Nonetheless, content experts quickly can learn the basics of item-writing.[36]

The term *test item* refers to "the basic unit of observation of any test."[37] Constructed response[38] and multiple-choice[39] are subcategories of test

---

[29] *Id.* at 137–38 ("Over all four examinations, the unfocused item stem (in combination with other flaws) and the negative stem (with AOTA) accounted for 67 flawed items (67%). The AOTA option, the NOTA option, and the partial-K type item accounted for an additional 23 flawed items.").

[30] *Id.* at 140.

[31] *Id.* at 141.

[32] *Id.*

[33] *Id.* at 142.

[34] HALADYNA, *supra* note 25, at 65 (citing Robert L. Thorndike, *The Analysis and Selection of Test Items*, *in* PROBLEMS IN HUMAN ASSESSMENT (Douglas N. Jackson & Samuel Messick eds., 1967)).

[35] *Id.*

[36] *Id.* at 187.

[37] *Id.* at 3.

[38] Constructed response items are those in which the test taker composes a response rather than selecting from provided responses. *See id.* at 49 (noting that short-answer items and essay are types of constructed response items).

items.   The typical law school multiple-choice item begins with a fact problem (called the root), followed by a question (called the stem) and then the answer choices (called the options).[40]   There are established guidelines for developing multiple-choice items.[41]   These guidelines address content, writing the root, writing the stem, and writing the options.[42]

A. Content

Each multiple-choice item should address one type of cognitive behavior and one type of content.[43]   For example, a multiple-choice item might address the application of knowledge (cognitive behavior) of the concept of tacking in adverse possession (content). Multiple-choice items should be based upon material that is important to learn, not on trivial content.[44] Content that is too specific risks being trivial.[45] Content that is too general risks being ambiguous.[46]   Questions "[s]hould involve representative and basic issues rather than unusual or marginal points which may be beyond the knowledge of most students" unless those points were "clearly focused on in the class or readings."[47] Multiple-choice items should be based on new scenarios, not scenarios already examined in class.[48]

Although the item-writer should strive to use new scenarios, the item-writer should avoid using "terms, cases, concepts or doctrines which are beyond the scope of the course."[49]   If the item-writer assumes student knowledge of even basic concepts from other first-year courses while

---

[39] *Id.* at 67–91 (explaining a variety of multiple-choice formats all of which require the test-taker to choose among provided responses).

[40] VOL. 2 JOSEPHSON, *supra* note 20, at 350.

[41] *See* HALADYNA, *supra* note 25, at 97–127 (providing detailed guidelines for developing multiple-choice items); VOL. 2 JOSEPHSON, *supra* note 20, at 369–80 (providing information on how to construct multiple-choice items for law school testing).

[42] *See* HALADYNA, *supra* note 25, at 99–100; VOL. 2 JOSEPHSON, *supra* note 20, at 369–80.

[43] HALADYNA, *supra* note 25, at 98–100.

[44] *Id.* at 100.

[45] *See id.* at 102.

[46] *See id.*

[47] VOL. 2 JOSEPHSON, *supra* note 20, at 373.

[48] *See* HALADYNA, *supra* note 25, at 101.

[49] VOL. 1 JOSEPHSON, *supra* note 15, at 13.

writing multiple-choice items for a first-year exam, this will result in flawed items.[50] For example, assuming that students understand the law of adverse possession in any first-year course other than Property would be problematic, unless that concept also had been taught in the course in question.

The item-writer should avoid writing items that call for student opinions or trick items. Trick items can result when "[t]he discrimination among options was too fine," "[m]ultiple correct answers were possible," "[p]rinciples were presented in ways that were not learned," or by other means.[51] For example, an item that requires students to choose among options on which experts would disagree would be a trick item. Most trick items likely are written inadvertently. No specific intent is required.

B. Writing the Fact Problem (The Root)

The facts in the root should be sufficient "to present the issues completely and fairly."[52] Thus, a well-written root would allow students to issue-spot early in the item. Inadequate facts may cause students to assume facts that could cloud the issues.[53] Therefore, while "[i]t is reasonable to include some irrelevant and decoy facts to test issue-spotting and other evaluation skills[,] . . . too many inessential details unfairly complicate the major issues being tested and should be avoided in a test with limited time."[54] Faculty should keep in mind that test items that seem simple to them as subject matter experts will almost always be more difficult for students.[55] Accordingly, "Fact situations should be realistic and plausible, the kind in which an attorney might conceivably be involved; highly artificial situations which are sometimes concocted for classroom use to dramatize a particular point should be avoided."[56]

---

[50] *Id.* ("A similar problem can arise with respect to the use in a test item of specialized vocabulary, including jargon not clearly within the student's knowledge.").

[51] HALADYNA, *supra* note 25, at 103–04 (citing Dennis M. Roberts, *An Empirical Study on the Nature of Trick Questions*, 30 J. EDUC. MEASUREMENT 331 (1993)).

[52] VOL. 2 JOSEPHSON*, supra* note 20, at 373.

[53] *See id.*

[54] *Id.*

[55] *See id.* at 370.

[56] *Id.* at 374.

In order to achieve clarity, distinct identities should be created for parties, places and things throughout the fact pattern.[57] The item writer should avoid creating confusion or ambiguity.[58] For example, while possibly amusing for the writer, alliterative (Jim, Jon, Jan) or rhyming (Mick, Dick, Rick) party names may increase confusion for students.

C. Writing the Question (The Stem)

The stem of a multiple-choice item should contain the "call of the question,"[59] and "[t]he test taker should always know what is being asked in the item after reading the stem."[60] An unfocused stem is one that "fails to provide adequate information to address the options."[61] For example, a stem that merely states "The court will" followed by several options would be an unfocused stem.[62] Contrast the previous example with "How will the court rule on a motion to suppress?" which would be a focused stem.[63] Multiple-choice items that contain no stem at all would be problematic because the test taker would have to attempt to determine what is being asked from reading the answer choices.[64] New facts can be added to the stem if done clearly.[65]

The item-writer should avoid negative wording in stems because negatively phrased items increase the strain on short-term memory.[66]

---

[57] *See id.*

[58] *Id.*

[59] *See id.* at 374; VOL. 1 JOSEPHSON, *supra* note 15, at 9 (defining the call of the questions as "a directive or interrogatory which specifies the task assigned to the student").

[60] HALADYNA, *supra* note 25, at 108.

[61] *Id.*; *see also* Downing, *supra* note 26, at 135 (noting that an unfocused stem is one which does not pose a direct question).

[62] *Cf.* HALADYNA, *supra* note 25, at 108 (discouraging stems that "might be too brief and uninformative to the test taker").

[63] *Cf. id.* (stating that the "test taker should always know what is being asked in the item after reading the stem").

[64] *See* VOL. 1 JOSEPHSON, *supra* note 15, at 9 ("All test items, whether they be objective, short answer, or essay, must contain a directive or interrogatory which specifies the task assigned to the student.").

[65] VOL. 2 JOSEPHSON, *supra* note 20, at 375.

[66] HALADYNA, *supra* note 25, at 111 (citing Pinchas Tamir, *Positive and Negative Multiple-Choice Items: How Different Are They?*, 19 STUD. EDUC. EVALUATION 311 (1993)).

Research supports that for items with high cognitive demand, "negatively phrased items [are] more difficult."[67] If negative terms are used, they should be in bold, capitalized or underlined.[68]

D. Writing the Answer Choices (The Options)

If the root and the stem have been written to conform to standard item-writing guidelines, the student should be able to anticipate the options after reading the root and the stem. The best number of options has been the subject of debate. Haladyna notes that item writers should "use as many choices as possible, but three seems to be a natural limit." [69] There may be some advantage to having more options but only if all of the options are "discriminating."[70] Josephson recommends four or five options.[71] MBE questions typically contain four options.[72]

Some item-writers write the options so that there is only one absolutely correct option and the other options are incorrect.[73] It is not uncommon, though, for law school multiple-choice items to ask students to select the best response, which calls for relative judgment and "does not require the examiner to draft options which are absolutely true/correct and false/incorrect."[74] If the stem calls for the best or worst option, the options should require the student to identify the best or worst option relative to the other options rather than to distinguish between true/correct and false/incorrect.[75]

The correct option is called the key option, and the other options are called distractors.[76] Item writers should strive to "make all distractors

---

[67] *Id.*

[68] *Id.*

[69] *Id.* at 112.

[70] *Id.*; *see also* Downing, *supra* note 26, at 140 ("Item discrimination describes how effectively the test item separates or differentiates between high ability and low ability students . . . [T]est items that highly discriminate are desirable.").

[71] VOL. 2 JOSEPHSON, *supra* note 20, at 375–76.

[72] Rachel Adams, *Multiple-Choice Item Writing: Art and Science*, B. EXAMINER, Feb. 1992, at 5, 8–9.

[73] VOL. 2 JOSEPHSON, *supra* note 20, at 356.

[74] *Id.* at 358.

[75] *Id.* at 378.

[76] *See id.* at 350.

plausible."[77] A plausible distractor will appear to be correct to a student who possesses only a "low degree of knowledge."[78] For that reason, "[i]n item-writing a good plausible distractor comes from a thorough understanding of common student errors."[79]

Moreover, "Options should be written so that they do not provide clues for less-knowledgeable but more test-wise examinees."[80] For example, these clues could include specific determiners, which usually suggest an incorrect answer because they are too extreme.[81] "Always," "never," "totally," "absolutely," and "completely" are examples of specific determiners.[82] Another example of a clue to a right answer are clang associations, which occur when the item writer inadvertently repeats a word in both the stem and the option.[83] A clang association in both the stem and the correct option can clue test-takers to the correct option.[84] However, including a clang association in an incorrect option can result in a trick item.[85] Additionally, clues can be created when the writer creates options that are not homogeneous as to grammar and/or content.[86] Blatantly absurd or ridiculous options are also clues.[87] Finally, item-writers should keep the length of the options as uniform as possible, because there is a common tendency to make the correct option the longest.[88]

There are additional considerations for writing the options. The key option should be randomly assigned to the option positions.[89] The options should be phrased in the positive.[90] The option "none of the above" should be used sparingly and is recommended only for "complex problems with a

---

[77] HALADYNA, *supra* note 25, at 120.

[78] *Id.*

[79] *Id.* at 121.

[80] Adams, *supra* note 72, at 10.

[81] HALADYNA, *supra* note 25, at 117.

[82] *Id.* at 117–18.

[83] *Id.* at 118.

[84] *Id*

[85] *Id.*

[86] *Id.*

[87] *Id.* at 119.

[88] *Id.* at 116.

[89] *Id.* at 113.

[90] *Id.* at 117.

quantitative answer."[91] The option "all of the above" should be avoided because it "seems to draw students into test-taking strategies more than directly testing for knowledge."[92]

VI. VALIDITY

Validity "refers to the appropriateness, meaningfulness, and usefulness of the specific inference made from test scores."[93] The validity of individual multiple-choice test items "contributes to the validation of test scores."[94] While it is not necessary to inquire into the validity of faculty-prepared tests at the intensive level required for standardized tests,[95] faculty should be aware of general guidelines for item and test validity.

The item-writing guidelines discussed above in Part V are relevant to item validity. An examination of individual items for conformity to item-writing guidelines would be one step in a formal item review evidencing validity.[96] Although formal item review might be impractical for faculty-prepared tests,[97] faculty attention to the guidelines is desirable because adherence to basic item-writing guidelines reduces the risk that students might fail a test as the result of flawed test items.[98]

As a further safeguard, adherence to overall guidelines regarding test content and reliability would reduce the risk that a test is invalid. Content validity requires "that the test must be capable of isolating and measuring the knowledge, skills and other matters it is supposed to measure."[99] This means that a test is valid if it measures students according to what they were supposed to learn in that course.[100] Ideally, what students were supposed to learn was articulated by the professor who designed the course in a set of instructional objectives.[101]

---

[91] *Id.*

[92] *Id.*

[93] Rachel Slaughter, Charles Kunce & Marcia Kuechenmeister, *Bar Examinations: Performance or Multiple Choice*, B. EXAMINER, Aug. 1994, at 7, 13.

[94] HALADYNA, *supra* note 25, at 18.

[95] *See* VOL. 1 JOSEPHSON, *supra* note 15, at 7.

[96] HALADYNA, *supra* note 25, at 187–88.

[97] *Id.* at 183.

[98] *See supra* note 32 and accompanying text.

[99] VOL. 1 JOSEPHSON, *supra* note 15, at 7.

[100] *Id.* at 8.

[101] *See id.*

In order for a test to be valid, it must also be reliable.[102] "Reliability refers to the consistency with which a test assesses whatever it is supposed to measure,"[103] so that the test must assess more than just a portion of what students were supposed to learn. Josephson believes that "most law school exams are at least partially invalid because very few teachers make any effort to systematically isolate all of their instructional objectives and, therefore, exams normally tend to measure only some of the objectives."[104] The final grade for a one-semester course, or even a two-semester course, may be based upon a single final exam. The vast amount of material addressed in these courses increases the likelihood that a test may not reliably measure what was taught.

The test writer should make sure that the test includes "an adequate and representative sampling of the material covered in the course."[105] A sample is adequate if it contains a sufficient number of questions.[106] A sample is representative if it includes the major instructional objectives of the course.[107] Josephson recommends that "the test ought to be constructed by direct reference to a checklist derived from the instructional objective[s], a course outline, class assignment sheets, or even the table of contents of the casebook."[108] Lack of reliability not only adversely affects the validity of tests, but also "creates hostility and suspicion"[109] among students. Students who have worked hard to master a vast amount of material for a course "want a comprehensive test."[110]

VII. APPEAL AND ANSWER JUSTIFICATION PROCEDURES

Multiple-choice items are frustrating for students because there is no opportunity for them to point out ambiguities or to explain their answer choices.[111] Students have little recourse against flawed multiple-choice items, and "[e]ven the most rigorously designed multiple[-]choice exams

---

[102] *Id.* at 15.

[103] *Id.*

[104] *Id.* at 8.

[105] *Id.* at 16.

[106] *See id.*

[107] *See id.*; VOL. 2 JOSEPHSON, *supra* note 20, at 369.

[108] VOL. 1 JOSEPHSON, *supra* note 15, at 17.

[109] *Id.* at 16.

[110] *Id.*

[111] *See id.* at 10.

will include some defective questions."[112] Defective questions, "[n]ot only reduce the validity of the exam as a whole, they also create anxiety, frustration and hostility in students who are in no mood or position to allow for teacher error (which they readily construe as incompetence)."[113] Faculty could relieve student frustration and simultaneously develop a valuable source of information about student understanding by adopting an appeal or answer justification procedure.

Josephson suggests that faculty establish an appeal procedure by which students are allowed to keep the multiple-choice items and immediately after the test are given the opportunity to file an appeal regarding any items the student believes were defective.[114] The faculty member would rule on each appeal, giving the student no credit, partial credit or full credit.[115] Josephson reports that there was strong student support for the appeal procedure he used.[116] Given the time involved in creating good multiple-choice items, many faculty would likely resist a procedure that allows students to keep the items.

Faculty who do not want to give their multiple-choice items to students but who are concerned about the effectiveness of their tests might consider implementing a technique called answer justification. Answer justification provides students with the option of converting multiple-choice items perceived by students to be "ambiguous or confusing" into short essay answers.[117] After selecting what the student believes to be the best option, the student may write a short justification of that choice.[118] If the student has selected the correct option, there is no reason to score the justification.[119] If the student has selected an incorrect option, she or he may receive full credit, partial credit or no credit for the justification.[120]

A study of seventeen different multiple-choice exam administrations to undergraduate psychology students revealed that students used answer

---

[112] VOL. 2 JOSEPHSON, *supra* note 20, at 381.

[113] *Id.*

[114] *Id.* at 382.

[115] *Id.*

[116] *Id.*

[117] David K. Dodd & Linda Leal, *Answer Justification: Removing the "Trick" from Multiple-Choice Questions*, 15 TEACHING PSYCHOL. 37, 37 (1988).

[118] *Id.*

[119] *Id.* at 38.

[120] *Id.* at 37.

justification on only 1% of the total number of multiple-choice items.[121] Professors did not have to score 67% of the justifications because students had selected the correct options.[122] On a typical fifty item multiple-choice exam administered to fifty students, answer justification added only about twenty minutes to the total grading time.[123] Student response to answer justification was very favorable, even among students who had not used the technique, because they were reassured by the knowledge that it was available.[124]

The authors of the study concluded that answer justification yields benefits for both professors and students because answer justification nearly eliminates student complaints about ambiguous items and helps professors identify problematic items.[125] It also eliminates the need to rescore tests or otherwise adjust class grades when multiple-choice items prove to be defective.[126] Answer justification allows students to improve their test scores, may increase their sense of control over the test, encourages them to carefully examine each answer option, and may improve their perceptions of their professors.[127]

A variation on the answer justification procedure is described in a study by Nield and Wintre who penalized students for "a bad explanation of a right answer."[128] While a penalty would likely reduce the frequency with which students wrote justifications, professors would have to read all of the justifications in order to enforce the penalty.[129] When there is no penalty, professors do not need to read the justifications written to accompany correct option choices.[130]

Yet another variation on answer justification is an informal in-class polling procedure during which students are given the opportunity to

---

[121] *Id.*

[122] *Id.*

[123] *Id.*

[124] *Id.*

[125] *Id.* at 37–38.

[126] *Id.*

[127] *Id.* at 38.

[128] Anthony F. Nield & Maxine Gallander Wintre, *Multiple-Choice Questions With an Option to Comment: Student Attitudes and Use*, 13 TEACHING PSYCHOL. 196, 197 (1986).

[129] Dodd & Leal, *supra* note 117, at 38.

[130] *Id.*

"provide alternative reasoning for their wrong answers."[131] Faculty might not wish to devote instruction time to such an exercise and might have concerns about the possible tone of such a discussion. However, Haladyna believes it would allow "[t]he expert judgment of classroom learners [to] be marshaled for exposing ambiguous or misleading items."[132]

Doctrinal faculty might be concerned that law students would use answer justification more frequently and aggressively than the undergraduate students who were the subject of the referenced studies, but Nield and Wintre concluded that after adopting the format allowing students to comment in writing on multiple-choice items, "much less grader and instructor time is taken in confrontations with students, and we find rapport between the students and teachers to be improved."[133] The additional time spent grading answer justifications would likely save the greater amount of time that might be spent in frustrating office visits from law students who wish to challenge multiple-choice items.

Answer justification would enhance multiple-choice testing with benefits for both students and faculty. It would provide students with an "important safeguard against poorly written [multiple-choice] items."[134] It also would reduce the possibility that some students who otherwise would have passed a test might fail due to flawed items by alerting faculty to problematic items.[135] Faculty could use answer justification to create a greater sense of "autonomy support"[136] for students by reducing students' sense of helplessness in the face of possibly flawed multiple-choice items.[137] Increasing students' feelings of autonomy can have positive

---

[131] HALADYNA, *supra* note 25, at 198.

[132] *Id.*

[133] Nield & Wintre, *supra* note 128, at 199.

[134] HALADYNA, *supra* note 25, at 197.

[135] *See supra* note 32 and accompanying text.

[136] Kennon M. Sheldon & Lawrence S. Krieger, *Understanding the Negative Effects of Legal Education on Law Students: A Longitudinal Test of Self-Determination Theory,* 33 PERSONALITY & SOC. PSYCHOL. BULL. 883, 884 (2007) (explaining that autonomy support has "three prototypical features": providing subordinates with as many choices as possible, providing explanations when no choice is possible, and demonstrating awareness of and appreciation for subordinates' points of view).

[137] *See id.* at 893 (noting that students perceived greater autonomy support where faculty "were felt to be more focused on student concerns and to provide more choices and more

effects on students' psychological well-being, final law school GPA, bar exam results and career motivation.[138]

Answer justification would produce the additional benefit of providing faculty with valuable feedback about student misunderstanding in the course. Without an answer justification procedure, faculty can only speculate about why students selected incorrect options. Answer justification would give the student the opportunity to explain the thought process that led to the incorrect choice. Moreover, "[E]vidence of misunderstanding is incredibly valuable to teachers, not a mere mistake to be corrected. It signifies an attempted and plausible but unsuccessful transfer [of understanding]. . . . For *attentive* students not to 'get it' is to show us that what we thought was clear was really not so."[139] Faculty could use this feedback to improve curriculum and teaching methods in order to achieve greater transfer of understanding in the future. Using tests to provide feedback not only from faculty to students about student performance but also from students back to faculty about student understanding would be one small step toward moving law school assessment to a higher level.[140]

VIII. CONCLUSION

Adherence to multiple-choice item-writing guidelines would yield more effective tests and enable academic support faculty to advise students more accurately about the item formats they are likely to encounter on tests. The use of appeal or answer justification procedures would alert faculty to possibly flawed multiple-choice items, reduce the amount of time spent in frustrating post-exam conferences with students, improve faculty/student relationships by reducing student anxiety, increase students' sense of autonomy, and provide faculty with valuable information about student misunderstanding.

---

meaningful rationales for mandatory rules and requirements" and less autonomy support where faculty were felt "to be more controlling and insensitive to their perspectives").

[138] *Id.*

[139] GRANT WIGGINS & JAY MCTIGHE, UNDERSTANDING BY DESIGN 51 (2d ed. 2005).

[140] *See* WILLIAM M. SULLIVAN ET AL., EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW 180 (2007) ("Putting these two kinds of assessment together—linking feedback *to* students with feedback *from* students about how well there are achieving the learning goals of a course—moves assessment to a higher level. We call this level institutional intentionality.").