Developing High-Quality Multiple-Choice Questions for Assessment in Legal Education

Susan M. Case and Beth E. Donahue

Multiple-choice questions have not been popular as assessment tools in law schools, where they are viewed as less intellectually rigorous than essay questions and less realistic in their relationship to the actual practice of law.[1] Lawyers write briefs and interpret cases and construct arguments. They do not select advice or arguments from a list of choices. Yet the field of study concerned with the science of educational testing, known as psychometrics, has amassed extensive evidence showing that multiple-choice questions have many advantages over other question formats.[2] This paper begins with a brief summary of the relative merits of multiple-choice questions as opposed to essay questions, and then focuses on how to construct multiple-choice questions to ensure that they will assess the competencies that are intended.

Decisions about Testing Format

After outlining the test's purpose and specifying its content, the assessor must choose the testing format. Although format alone does not guarantee

Susan M. Case has been the Director of Testing for the National Conference of Bar Examiners since 2001.

Beth E. Donahue is the MBE Program Director at the National Conference of Bar Examiners.

1. While multiple-choice questions on the Multistate Bar Examination form the backbone of most U.S. jurisdictions' bar exams, "[w]ithin...legal education...[multiple-choice questions] still have something of a bad reputation. Yes, they are easy to mark, but you can't test more than surface learning of facts...[and] an averagely lucky monkey could expect to get reasonable marks in a[multiple-choice] test." Edwina Higgins and Laura Tatham, Exploring the Potential of Multiple-Choice Questions in Assessment, Learning and Teaching in Action Issue 1 (Winter 2003), available at <http://www.ltu.mmu.ac.uk/ltia/issue4/higginstatham.shtml> (last visited Nov. 12, 2008).

2. Much of the research on testing in the professions has been done in the field of medicine. Most, if not all, medical schools have an office of medical education that conducts such research. In addition, substantial research has been completed at the National Board of Medical Examiners, the nonprofit organization that develops the United States Medical Licensing Examination. Over time, the medical licensing examination has included essay tests, oral examinations, standardized patient examinations, computer-based simulations, and multiple-choice tests. Comparison of these formats has been the focus of significant research.

that a test is of high quality or that it assesses higher-order skills, there are several clear advantages and disadvantages of well-crafted questions in every format. A comparison of essay and multiple-choice formats reveals that: (1) essay tests are relatively easy to develop, but difficult to grade, particularly in maintaining consistent grading standards across a large number of papers; and (2) essay tests assess writing and in-depth analytical skills but have limited content coverage, advantaging those who are lucky in the selection of topics.

In contrast, multiple-choice questions are relatively difficult to develop, but easier to grade, inherently ensuring consistent standards across any number of examinees. Multiple-choice tests do not assess writing skills, nor do they assess in-depth knowledge of a particular topic. Because they assess a large number of topics, they reduce the likelihood that someone will be lucky or unlucky in the selection of topics.

Poorly written questions may test nitpicky facts; they may contain ambiguous language; they may contain unrealistic scenarios. These flaws are seen across testing formats; they are construction-related concerns rather than inherent drawbacks associated uniquely with multiple-choice questions.[3] Poorly crafted exam questions, whether they be essays, multiple-choice questions, or oral interview questions, fail to assess what is intended and fail to encourage the learning that is desired. In other words, the particular question format or combination of formats is not as important as the skill and sophistication with which the questions are crafted and the exam as a whole is assembled and graded.[4]

The format selected for a test should be tied to the knowledge and skills one hopes to measure.[5] An oral examination generally is inappropriate to assess reading skill. Similarly, a multiple-choice exam generally is inappropriate to measure writing skill. But where the goal is to measure an examinee's knowledge of a particular field or to measure an examinee's ability to apply legal reasoning to a variety of fact patterns, the multiple-choice format has several distinct advantages over other formats, most notably in content coverage, grading ease, grading

3. See Cynthia B. Schmeiser and Catherine J. Welch, Test Development, *in* Educational Measurement 318 (Robert L. Brennan, ed., 4th ed., Westport, Conn., 2006) (summarizing research showing that both multiple-choice questions and constructed-response questions, properly drafted, can "measure higher-order thinking skills").

4. See Susan M. Case, Assessment Truths We Hold as Self-Evident and Their Implications, Advances in Medical Education 2-6 (Dordrecht, 1997). See also Standards for Educational and Psychological Testing 3 (Washington, D.C., 1999) ("Tests differ on a number of dimensions: the mode in which test materials are presented....the degree to which stimulus materials are standardized; the type of response format (selection of a response from a set of alternatives as opposed to the production of a response); and the degree to which test materials are designed to reflect or simulate a particular context. In all cases, however, tests standardize the process by which test-taker responses to test materials are evaluated and scored [, and] the same general types of information are needed for all varieties of tests.").

5. See Schmeiser and Welch, Test Development, *supra* note 3, at 318; Standards for Educational and Psychological Testing, *supra* note 4, at 38.

consistency, and reliability of scores.[6] Research has consistently demonstrated a high correlation between scores on essays and multiple-choice questions: examinees who do well on a well-crafted essay exam in a particular field also tend to do well on a well-crafted multiple-choice exam in that field, and those who do poorly on essays likewise tend to do poorly on multiple-choice questions.[7] The correlation is not perfect; essays do provide some additional information (e.g., level of writing ability) that may be helpful to law professors in holistic assessments of their students.

When testing format considerations lead to the selection of the multiple-choice format as a component of a test, the assessor must invest a significant amount of time in drafting and editing the multiple-choice questions to be used. How, then, can an instructor draft high-quality multiple-choice questions?

Drafting Multiple-Choice Questions

High-quality multiple-choice questions contain three components: the *stem* (a scenario or vignette setting up the question), the *lead-in* (the question), and the *options* (answer choices, typically labeled A, B, C, etc.). While the three components can be written in any order, they will be discussed below in the order in which they are ultimately presented to examinees. Following the discussion of question components is a set of edited questions showing "before" and "after" versions in a side-by-side format, as well as a summary checklist for writing multiple-choice questions.

6. Reliability in the context of educational measurement refers to the precision of scores, or "how the scores resulting from a measurement procedure would be expected to vary across replications of that procedure." Edward H. Haertel, Reliability, *in* Educational Measurement, *supra* note 3, at 65. When a single exam will in large part determine examinees' grades for an entire semester, as is common in law school, the reliability of that exam should be of significant concern to the assessor.

   One study that compared essays to multiple-choice questions for the amount of testing time needed to achieve a target level of reliability (a level of 0.80) found that essays required double to triple the amount of testing time, depending on how they were graded. David Swanson and Susan Case, Trends in Written Assessment: A Strangely Biased Perspective, Approaches to the Assessment of Clinical Competence 38-53 (R. Harden, I. Hart, and H. Mulholland, eds., Norwich, England, 1992).

7. See Swanson and Case, Trends in Written Assessment, *supra* note 6, at 38-53; Schmeiser and Welch, Test Development, *supra* note 3, at 318 (scores from constructed-response questions "highly correlated" to those from multiple-choice questions when used to "measure the same domain of knowledge and skills"). Significant correlations have likewise been found between the Multistate Essay Examination, the Multistate Performance Test, and the Multistate Bar Examination, which are designed to test overlapping (but not identical) areas of legal knowledge and skill.

Figure 1
The Components of Multiple-Choice Questions

Stem (or Scenario or Vignette)	A seller and a buyer entered into a contract obligating the seller to convey title to a parcel of land to the buyer for $100,000. The agreement provided that the buyer's obligation to purchase the parcel was expressly conditioned upon the buyer's obtaining a loan at an interest rate no higher than 10 percent. The buyer was unable to do so, but did obtain a loan at an interest rate of 10.5 percent and timely tendered the purchase price. Because the value of the land had increased since the time of contracting, the seller refused to perform. The buyer sued the seller.
Lead-in	Will the buyer prevail?
Options (including the correct answer, or key,* and the incorrect answers, or distractors)	(A) No, because an express condition will only be excused to avoid forfeiture. (B) No, because the contract cannot be modified without the consent of the seller. (C) Yes, because the buyer detrimentally changed position in reliance on the seller's promise to convey. (D) Yes, because the buyer's obtaining a loan at an interest rate no higher than 10 percent was not a condition to the seller's duty to perform.*

The Stem

When drafting a stem, the primary consideration is that the content be relevant to the assessor's purpose in giving the exam in the first place. For example, in the case of the Multistate Bar Examination, the purpose of the exam is to assess the examinees' readiness, in terms of applying fundamental legal principles to fact scenarios, for the beginning practice of law. The following MBE question uses a straightforward stem to test an important concept in product liability:

* The key for each sample multiple-choice question in this paper is designated by an asterisk (*).

A restaurant hired an exterminator to eliminate cockroaches from the basement under the restaurant. Around midnight, the exterminator applied to the basement floor and walls an effective pesticide that he had purchased from the manufacturer. A toxic gas released by the pesticide penetrated into the restaurant kitchen and did not disperse by the next day. As a result, the restaurant was required to close that day.

The restaurant brought a tort action based on product liability against the pesticide manufacturer for lost profits.

Will the restaurant prevail?

(A) No, because in this action purely economic loss is not recoverable.*

(B) No, because the exterminator was the proximate cause of the restaurant's damages.

(C) Yes, because the manufacture of pesticides is an abnormally dangerous activity.

(D) Yes, because the pesticide was being used as intended.

It looks simple, but anyone who has attempted to write multiple-choice questions knows that writing a concise stem that adequately sets up a legal issue (supports the intended correct answer), without giving too much away or accidentally supporting an unintended option, is a difficult art. Unfortunately, we cannot recommend any shortcuts. Much of the multiple-choice stigma–a stigma that is deserved for poorly crafted questions–comes from multiple-choice questions that *have no vignettes*. For example,

Purely economic recovery will be barred in which of the following causes of action?
(A) Negligence.
(B) Fraud.
(C) Defamation.
(D) Product liability.

Or, worse yet,

Purely economic loss is recoverable in a product liability action.
(A) True.
(B) False.

Questions that contain no vignettes, merely asking examinees "What's the rule?" or "What's true?", are testing surface learning–the examinees' recall of isolated facts ("recall questions"). Recall questions can often be answered by turning to a single paragraph in a textbook. They reward examinees who have simply memorized the material, but who might not be able to apply or interpret it.[8]

One common problem with question stems is a failure to test important content. Questions that test obscure legal principles that are not likely to be addressed by new lawyers, or those that use fact scenarios that are not likely to occur in the real world, simply don't pass the "who cares?" test and do not further the purpose of a high-stakes exam. For example,

> Deon went to his girlfriend Clara's apartment for her birthday party. He secretly hung a new $300 camel-hair coat in Clara's closet, then pressed a note into her hand that read, "If you look in your closet, you'll see your birthday present." Clara, busy mixing drinks for her guests, stuffed his note into her pocket without reading it. Later, she noticed the coat and admired it. She assumed it belonged to a guest.
>
> Deon became ill and left the party early. Shortly thereafter, Clara's friend Margaret arrived. Remembering that Margaret needed a winter coat, Clara pointed out the camel-hair coat in the closet and suggested that Margaret steal it. After much persuasion, Margaret agreed to do so. She took the coat home, where she found a card in the pocket that read, "Happy Birthday, Clara! Love, Deon."
>
> Margaret called Clara and told her. Clara laughed and said, "Yeah, Deon left it there for my birthday present. He called to see if I'd found it, and just about died when I told him you'd taken it. I guess you'd better bring it back." Margaret did so the next day.
>
> In this common-law jurisdiction, larceny of an item worth $50 or more is a felony.

8. See Susan M. Case, David B. Swanson, and Douglas R. Ripkey, Verbosity, Window Dressing, and Red Herrings: Do They Make a Better Test Item?, 71 Academic Medicine 10, S28, S30 (Oct. Supp. 1996). Questions drafted in the true/false format, if they are to avoid ambiguities, frequently fall into the realm of recall questions. See Susan M. Case and David B. Swanson, Constructing Written Test Questions for the Basic and Clinical Sciences 18 (3rd ed., Philadelphia, 2001). For this reason, true/false questions have been eliminated from both the medical and legal licensure exams. To understand why recall questions are generally inappropriate to assess higher-order skills, it helps to consider what the ultimate challenges will be for practitioners in the field of interest. As one study of medical students' diagnostic errors determined, the most common student shortcoming was not in knowledge, which presumably can be assessed by recall questions, but in analytical reasoning, which cannot be assessed by recall questions. In diagnosing standardized patients, the medical students most commonly erred by (1) basing their conclusions on accurate but nondiscriminating information, and (2) making incorrect inferences from accurate and discriminating information. See Michael H. Friedman et al., Medical Student Errors in Making a Diagnosis, 73 Academic Medicine 10 at S19, S21 (Oct. Supp. 1998).

Clara may be convicted of
(A) larceny and conspiracy.
(B) conspiracy and solicitation.
(C) conspiracy or solicitation, but not both.*
(D) larceny, conspiracy, and solicitation.

This question has several flaws, the most fundamental of which is the content laid out in the stem. Is Clara likely to be prosecuted? If not, asking what she "may be" convicted of is not a sufficiently important issue to be tested. Such questions not only give examinees a negative impression of the exam (that they're being tested on things that they'll never need to know to practice law), but they also can undermine the validity of scores if the assessment purports to measure more than esoteric knowledge.[9]

A stem is typically drafted with a particular legal issue in mind. Once a reasonable stem has taken shape that sets up an issue that is pertinent to what the assessor is trying to test, the next step is to review the stem carefully to eliminate any technical flaws that might undermine the question's performance on an exam.[10] We recommend the following editorial guidelines:

- Use concise, standard English (avoiding legalese).
- Use the minimum number of actors needed to test the legal concept at issue.
- Use common descriptions for actors rather than names or functional names ("the attorney" rather than "Attorney," "Alpha," or "Albert").
- Avoid sensitive and sensational topics whenever possible; if such topics must be covered, write the stems in a manner that is as minimally salacious as possible.
- Eliminate any subtle gender, age, ideological, racial/ethnic, or rural/urban biases.
- Eliminate any "red herrings" or other facts intended to trick examinees.

The first guideline above is worth emphasizing. Remember that a key advantage of the multiple-choice format is that it allows the assessor to test a multitude of issues in a small amount of testing time. Long, confusing questions reduce that advantage. For example, in the coat-theft question

9. Validity in the context of educational measurement refers to "the extent to which the evidence supports or refutes the proposed interpretations and uses" of test scores. See Michael T. Kane, Validation, *in* Educational Measurement, *supra* note 3, at 17). Put another way, validity refers to the extent to which test scores accurately estimate examinees' competencies in the domain of interest. See Case and Swanson, Constructing Written Test Questions for the Basic and Clinical Sciences, *supra* note 8, at 10-11.

10. A question's "performance" for purposes of this article refers to the question's difficulty (a simple statistic showing the percentage of examinees who answered the question correctly) and the question's discrimination (a statistic showing how well the question distinguished between examinees who performed well on the exam as a whole and those who performed poorly). Performance statistics are easy to calculate for multiple-choice exams and can be useful in reviewing the quality of individual questions.

above, notice how many extraneous facts are included in the stem: Clara admires the coat, Deon becomes ill, Margaret needs persuasion, Clara and Margaret laugh about the whole thing, and so on.

Multiple-choice stems should set up simple, understandable problems; the challenge for examinees should be in determining the correct answer, not in trying to understand the scenario. As discussed more fully below, the difficulty of a multiple-choice question is best determined by the sophistication and plausibility of the options, not by the complexity of the stem.

The Lead-in

The lead-in is the component of the multiple-choice question that frames the examinees' task. For the purposes of this article, the salient issue in drafting a lead-in relates to the specificity of the question asked. We recommend focused questions that pose a single, specific task and that naturally flow into a set of homogeneous options.

Consider the following lead-ins:

Given the above facts, which of the following statements is true?

Given the above facts, of which crime is the defendant guilty?

Note that the first lead-in is not focused. It might easily lead to a set of heterogeneous options—an "apples to oranges" problem for examinees. In contrast, the second lead-in is so specific that a knowledgeable examinee could answer it without options from which to choose.[11] This level of specificity is the goal when drafting a lead-in.

Each multiple-choice lead-in controls the task demanded of the examinees in answering that question, hopefully assessing a particular skill that is deemed desirable. As discussed above, multiple-choice questions generally are not useful in assessing writing skill; to pose a writing task the assessor obviously must use another format. Yet a wide range of tasks is possible and appropriate in a multiple-choice exam. In particular, multiple-choice questions can require lawyering tasks, as in "What question should the lawyer pose next?" They can assess lawyering strategy, as in "What is the best argument against enforcement of the contract?" The following lead-ins, loosely sorted by tasks, have appeared in recent administrations of the Multistate Bar Examination:

Gathering Information

- What issue must the lawyer resolve before advising his client?
- What information does the lawyer need to proceed?
- What question should the lawyer ask next?
- Which source will be the most helpful?

11. Note that a well-crafted stem and lead-in of a multiple-choice question are no different from those of a short-answer question. The only difference to the drafter is the addition of the options.

Identifying dispositive issues/Formulating strategy

- What is the best defense?
- Whom should the worker sue?
- What is the biggest obstacle to relief?
- What is the appropriate burden of persuasion?
- What test must the judge apply?
- What is the strongest objection?
- Which law governs?

Synthesizing law and facts/Predicting outcomes

- Is the ordinance constitutional?
- Should the owner collect damages?
- Should the conviction be overturned?
- Should the statement be excluded?
- Is the opinion admissible?
- How should the proceeds be distributed?
- Is judicial notice proper?
- What are the boundaries of the owner's land?
- Who owns what interest in the land?
- Who will prevail?

We recommend the following editorial guidelines with respect to lead-ins:

- Avoid sentence fragments (e.g., "The defendant is guilty of:").
- Avoid negative framing (e.g., "Which of the following arguments is the LEAST helpful to the plaintiff?").

The Options

The correct option is called the answer *key*, while the incorrect options are called *distractors*. There is no rule restricting the number of options to the traditional four or five; in fact, research suggests that more options (so long as they remain plausible to some segment of the examinees) will improve the performance of questions, perhaps by reducing the occurrence of truly random guessing.[12] What is most important substantively is that the options, in whatever number, be concise, parallel, and plainly responsive to the lead-in. Two good examples follow:

12. See Susan M. Case, David B. Swanson, and Douglas R. Ripkey, Comparison of Items in Five-Option and Extended-Matching Formats for Assessment of Diagnostic Skills, 69 Academic Medicine 10 at S1 (October Supp. 1994) (studying reliability differences by question format on the U.S. Medical Licensing Examination, Step 2, and finding higher reliability in questions with 9-23 options than in questions with 5 options).

Which of the following causes of action is most likely to be successful?
(A) Battery.
(B) False imprisonment.
(C) Libel.
(D) Trespass.

Who will likely prevail?

(A) The landlord, because the condemnation superseded and canceled the lease.
(B) The landlord, because the parties specifically agreed to the consequences of condemnation.
(C) The tenant, because the landlord breached his implied warranty of quiet enjoyment.
(D) The tenant, because otherwise the landlord would be unjustly enriched.

The performance of a multiple-choice question may be sabotaged by options that add new twists to the lead-in or that involve multiple layers of analysis. One of the most problematic examples is known as the "K-type" question,[13] or the "multiple true/false" question. An example of a K-type question follows:

Which of the following would be proper for an attorney to consider in determining the basis for hourly fees after an initial consultation?

I. The novelty and difficulty of the questions involved in a particular case.
II. The fee customarily charged in the locality for similar legal services.
III. Whether the client is a lawyer or a member of the attorney's immediate family.
IV. The length of the initial consultation.

(A) I and II, but not III and IV.
(B) II and III, but not I and IV.
(C) I, II, and III, but not IV.
(D) I, II, III, and IV.

K-type questions essentially have two sets of options, the second of which refers back to the first. These questions do not generally perform as

13. There is no intrinsic meaning in the label "K-type." Multiple-choice formats have been named alphabetically by the measurement community in the order invented.

well as traditional multiple-choice questions,[14] perhaps because it is easy, under stress, for even the most knowledgeable examinees to get tangled up in the extra analysis required to answer them. It is also worth noting that the complexity is not related to legal reasoning, but to a more general error in question construction.

Once a set of options is constructed to respond to the lead-in, it is important to review the options carefully to ensure that the intended key is clearly the best answer. Frequently, additional editing to the stem or to the options will be necessary.[15] The options are the assessor's main tool in controlling a multiple-choice question's difficulty. As an illustration, consider the difference in difficulty between the two sets of options proposed below (ignoring the fact that the question is a "recall question" that does not test higher learning):

> Which of the following is the largest city?
>
> Option Set #1 makes this an easy question:
> (A) Amesville
> (B) Bridgeport
> (C) Chicago
> (D) Davenport
> (E) Omaha
>
> Option Set #2 makes this a difficult question:
> (A) Austin
> (B) Baltimore
> (C) Chicago
> (D) Dallas
> (E) Philadelphia

Ideally, each option in a multiple-choice question will be selected by a subset of examinees. More plausible distractors will draw more examinees, resulting in a more difficult question. In the case above, the distractors in Option Set #2 are all more plausible answers to an examinee with limited knowledge than the distractors in Option Set #1. Plausibility can be achieved by drawing on common misconceptions and faulty reasoning. It is not related to the sophistication of the language used, the length of the question, or the number of issues and nonissues presented, all of which add *irrelevant* difficulty. Irrelevant difficulty means that more examinees will

14. See Case and Swanson, Constructing Written Test Questions for the Basic and Clinical Sciences, *supra* note 8, at 122 (summarizing findings at the National Board of Medical Examiners that K-types were less discriminating and less efficient than traditional multiple-choice questions).

15. If the options are drafted after the stem, the drafter should review the stem for superfluous information at this stage as well. To avoid red herrings in the stem that do not tie into at least one option choice, try drafting the options *before* the stem.

answer the question incorrectly, but not in a pattern that is related to the examinees' mastery of the subject area. We recommend the following editorial guidelines with respect to issues of difficulty:

- Avoid long or complicated options.
- Avoid vague frequency terms such as *rarely* and *usually*; also avoid absolute terms such as *always* and *never*.
- Use parallel language across all options, and make sure all options are of similar length and complexity.
- Present options in a logical order (or in alphabetical order, if there is no logical order).
- Avoid using "none of the above" or "all of the above" as options.
- Avoid conditioning the options on information not presented in the stem (e.g., "Yes, if the jury finds that..."; "No, unless the seller also offered...").

Sample Multiple-Choice Question Edits

The following questions demonstrate how common pitfalls in writing multiple-choice questions can be resolved through careful editing.

Sample Question #1: Length and Complexity. Note how the edited version saves time while testing the same legal principle.

Original	Edited Version
Adam and Bruce are tenured professors at State University. Recently, they drafted and played key roles in the adoption by university officials of a new campus speech code prohibiting certain kinds of speech on the campus. Students, staff, and faculty convicted by campus tribunals of violating the speech code were made subject to penalties that include fines, suspensions, expulsions, and termination of employment. There was widespread public opposition to the campus speech code and a great deal of public anger directed toward Adam and Bruce because of their role in its drafting and adoption. When the appropriation bill for State University was considered by the state legislature during the following year, critics of the speech code were successful in cutting the university's budget and in including a severable provision in the appropriation statute declaring that "none	A tenured professor at a state university drafted a new university regulation prohibiting certain kinds of speech on campus. The regulation was widely unpopular. The following year, the state legislature approved a severable provision in the appropriations bill for the university declaring that none of the university's funding could be used to pay the professor, who was specifically named in the provision. In the past, the professor's salary had always been paid from funds appropriated to the university by the legislature, and the university had no other funds that could be used to pay him. Is the provision constitutional? (A) No, because it was based on conduct the professor engaged in before it was enacted.

Question 1 (con't).

Original	Edited
of the funds appropriated herein shall be paid to Professor Adam or Professor Bruce." In the past, all of the salaries of these two professors were paid from funds provided by the annual state appropriation for the university; the university has no other funds with which to pay them. In a suit brought in an appropriate state court by Adam and Bruce against the state officials who will determine whether they are paid from this appropriation, Adam and Bruce contend that the provision of the appropriation statute applying to them is unconstitutional. The court in that case will probably hold this provision of that statute (A) unconstitutional, because this prohibition on payments to Adam and Bruce was based on conduct they engaged in before the enactment of the appropriation statute and, therefore, violates the privileges and immunities clause of the Fourteenth Amendment. (B) unconstitutional, because the prohibition in the appropriation statute of any payment to Adam and Bruce amounts to the imposition of a punishment by the legislature without trial and, therefore, violates the bill of attainder clause.* (C) constitutional, because the full faith and credit clause requires all courts to enforce appropriation statutes enacted by state legislatures strictly according to the terms of those statutes. (D) constitutional, because the Eleventh Amendment prohibits any suit against a state that would interfere with its plenary power to appropriate and spend state funds in the manner that the state legislature deems most conducive to the welfare of its people.	(B) No, because it amounts to the imposition of a punishment by the legislature without trial.* (C) Yes, because the full faith and credit clause requires the court to enforce the provision strictly according to its terms. (D) Yes, because the state legislature has plenary power to appropriate state funds in the manner that it deems most conducive to the welfare of its people.

Sample Question #2: Specific, Focused Lead-in. Note how the edited version uses a more specific lead-in to focus the options and force them into a parallel form.

Original	Edited
Developer acquired a large tract of land. At the time of purchase, she had not decided what to do with the land; however, she felt that it had great potential and would increase in value. Approximately five years later, Developer decided to develop the tract into one-acre lots on which houses would be built, as permitted by the zoning ordinance in effect when she purchased the land. She hired an architect to prepare the plans. The plans were prepared and given to Developer. Developer then went to the local municipal building for the first time to discuss the plans with the appropriate officials. To Developer's surprise, she learned that the municipality had rezoned the tract of land two years earlier. The new zoning code imposed a five-acre minimum lot size for new home construction. The new zoning was enacted in compliance with required procedures and is appropriate to the area to which it applies. Developer asked her attorney whether she can proceed with her plans. Under these circumstances, the attorney should advise Developer that she (A) can proceed, using the zoning in effect at the time of purchase, because her rights to utilize the land vested at that time. (B) can proceed, using the zoning in effect at the time of purchase, if, but only if, she can prove that she purchased the property in reliance on such zoning. (C) cannot proceed, because the remedy for improper land use regulation is to seek inverse condemnation. (D) cannot proceed, and she is not entitled to any reimbursement from the municipality.*	A developer acquired a large tract of land with the intent to develop it for residential use at a later date. Five years later, she paid an architect to prepare plans for the land's development as a residential community of one-acre lots. The developer then discussed the plans with the appropriate municipal officials, who told her that the municipality had rezoned the tract of land two years earlier. The new zoning code imposed a five-acre minimum lot size for new home construction. The new zoning was enacted in compliance with required procedures and is appropriate to the area to which it applies. Can the developer proceed with her plans? (A) Yes, because her rights to utilize the land vested at the time she bought the land. (B) Yes, because she bought the property in reliance on the zoning in effect at that time. (C) No, because the remedy for improper land use regulation is to seek inverse condemnation. (D) No, because the new zoning restrictions are properly applied to the developer's current plans.*

Sample Question #3: Layers of Analysis. Note how the edited version cuts the additional layers of analysis required by the original options.

Original	Edited
On April 1, Painter contracted with Owner to paint Owner's office building beginning on June 1 for $20,000. Painter expected to finish the painting in one month. On May 1, Painter called Owner and stated that he was over-committed for June and was not sure that he could perform the painting on schedule, but that he would try. On May 8, Owner learned from his banker that Painter had had to fire part of his work crew because of financial difficulties. On May 15, Painter called Owner and stated that he would perform the painting contract only if Owner immediately promised to pay $30,000 for the work. Which of the following facts would support a cancellation of the contract by Owner? I. Painter's May 1 statement expressing doubt about meeting the work schedule. II. Owner's May 8 discovery of Painter's financial difficulties. III. Painter's May 15 demand from Owner for additional pay. (A) I only. (B) II and III. (C) III only.* (D) I, II, and III.	On April 1, a painter contracted with a business owner to paint the owner's office building beginning on June 1 for $20,000. The painter expected to finish the job in one month. On May 1, the painter called the owner and stated that he was over-committed for June and was not sure that he could perform the painting on schedule, but that he would try. On May 8, the owner learned from his banker that the painter had had to fire part of his work crew because of financial difficulties. On May 15, the painter called the owner and stated that he would perform the painting contract only if the owner immediately promised to pay $30,000 for the work. Is the owner entitled to cancel the contract as of May 15? (A) Yes, because of the doubts expressed by the painter about meeting the work schedule. (B) Yes, because of the painter's financial difficulties. (C) Yes, because of the painter's demand for additional pay.* (D) No, because the agreed-upon time for performance has not yet arrived.

Conclusion

Assessment is an important component of instruction. In addition to the obvious purpose of testing as a mechanism to determine grades, tests can also communicate to students what material is important, motivate students to study, identify areas of deficiency in need of remediation or further learning, and identify areas where the curriculum is weak.

A good test must satisfy two basic criteria. Each test question must first address important content, regardless of whether the test is a performance test, an essay test, a short-answer test, or a multiple-choice test. While the requirement that a test question assess important content is an essential

condition, it is not sufficient to guarantee that the question is a good one. Each test question must also be carefully constructed to avoid flaws that could reduce its effectiveness in testing the desired content. Assessors must be mindful of both content- and construction-related issues, as outlined in this paper, to ensure the high quality of the multiple-choice questions used in their examinations.

Appendix

Checklist for Writing Multiple-Choice Questions

Content

- Focus on important concepts; don't waste time testing trivial facts.
- Focus on application of knowledge, not recall of isolated facts.

The Stem

- Write concisely in standard English (avoiding legalese and proper nouns).
- Use the minimum number of actors needed to test the legal concept being addressed.
- Avoid superfluous information unless it is relevant to the legal complexity being assessed.
- Avoid "tricky" and overly complex facts.

The Lead-in

- Make sure the lead-in poses a clear question.
- Make the lead-in as specific as possible.
- Avoid negatively phrased lead-ins (e.g., "Which of the following is *not* . . .").

The Options

- Write as concisely as possible (one-word options are perfectly fine).
- List options in logical or alphabetical order.
- Make sure distractors are plausible.
- Make sure all options are similar in length and complexity.
- Avoid using absolutes such as *always*, *never*, and *all,* as well as vague terms such as *usually* and *frequently*.
- Avoid using conditions within the options (e.g., "Yes, but only if..."; "No, unless...").
- Avoid extra layers of analysis (e.g., "A and B are correct, but C is incorrect").
- Avoid any variation of "All of the above" or "None of the above."