

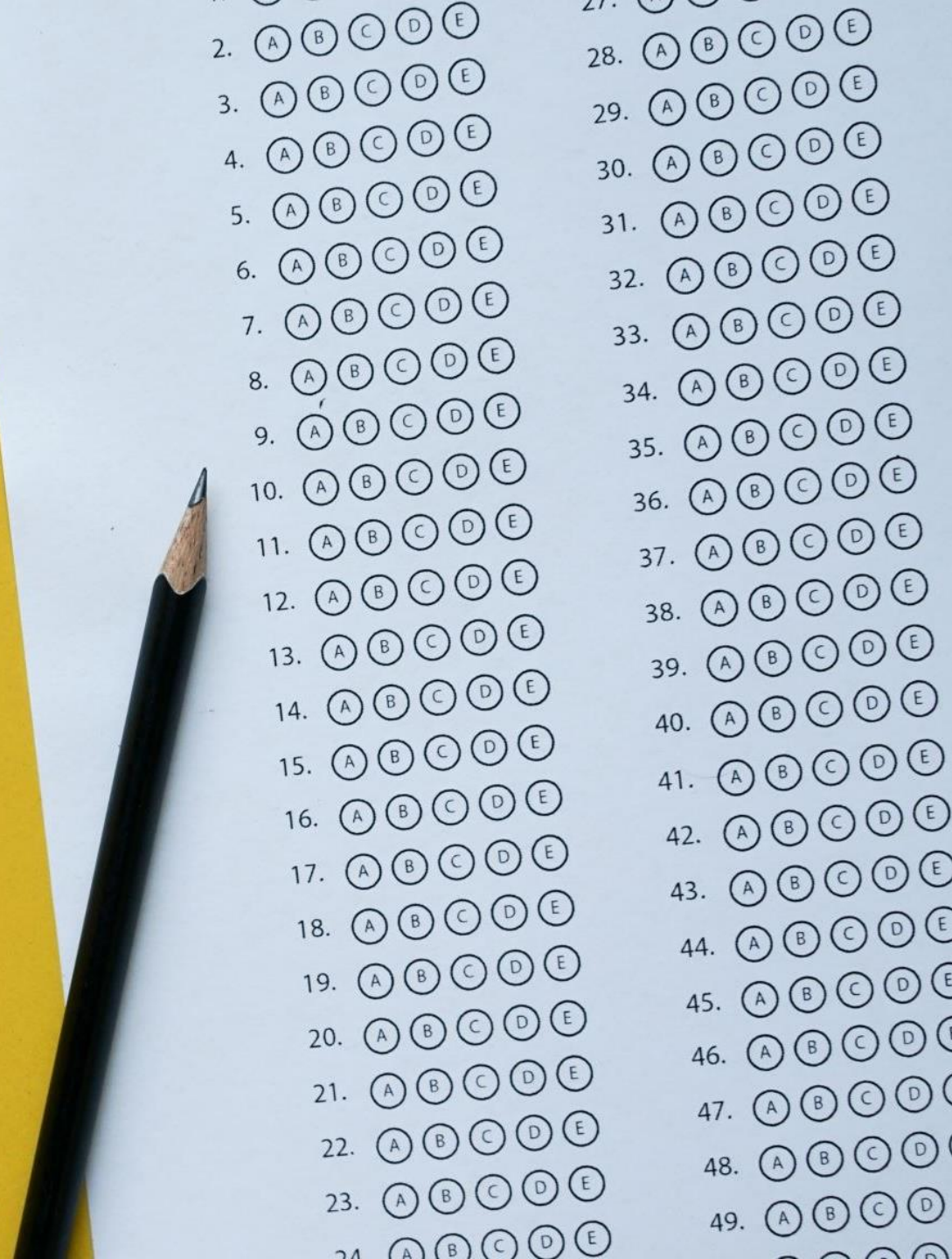
Writing Test Questions: CVM students in pre-clinical coursework

Elizabeth Davis, DVM, PhD, DACVIM
Interim Dean and Professor
Kansas State University, College of
Veterinary Medicine



Introduction

- Write learner-centered, measurable objectives.
- Write well developed knowledge/application-based response questions.
- Identify commonly made mistakes in writing MCQs.
- Improve questions in order to make scoring reliable and clear.



Elements of assessment

- **Validity** – how well does the assessment carry out the intent
- **Reliability** – consistency of measurement
 - If we were to make the same assessment again, would the outcome be repeated?
- **Defensibility** – appropriate standards of assessment
 - Legal implications of assessment instrument and outcome

Instructional Methods

**Validity, Reliability, and Defensibility of
Assessments in Veterinary Education**

Kent Hecker ■ Claudio Violato

JVME 36(3) © 2009 AAVMC

Reliability

- Make every effort to reduce error
- Characteristics of the instrument
 - Item types, item quality, difficulty
 - More items
 - Varying degrees of difficulty
- Characteristics of the sample
 - Participants interaction with the instrument
- Conditions of the administration
 - Testing conditions

VETERINARY EDUCATORS

Understanding Reliability: A Review for Veterinary Educators

Kenneth D. Royal ■ Kent G. Hecker



Accreditation policies and procedures of the AVMA Council on Education (COE)

2. PRINCIPLES OF ACCREDITATION

2.1 Standards of Accreditation

2.1.1 Standards of Accreditation

Standard 1 – Organization

Standard 2 - Finances

Standard 3 - Physical Facilities and Equipment

Standard 4 - Clinical Resources

Standard 5 - Information Resources

Standard 6 - Students

Standard 7 - Admission

Standard 8 - Faculty

Standard 9 - Curriculum

Standard 10 - Research Programs

Standard 11 - Outcomes Assessment

2.1.2 Standard Development

[AVMA COE POA](#)

Standard 9 - Curriculum

The curriculum must provide at least 130 weeks of direct instruction. The summative, concluding period of clinical instruction must include a minimum of 40 weeks of hands-on clinical education involving the prevention, diagnosis, treatment, or mitigation of disease related to animal health, or other experiential, workplace-based learning that is supervised through real-time interactions with the instructor(s). The curriculum and educational process should initiate and promote lifelong learning in each professional degree candidate.

The curriculum in veterinary medicine is the purview of the faculty of each college, but must be managed centrally based upon the mission and resources of the college. There must be sufficient flexibility in curriculum planning and management to facilitate timely revisions in response to emerging issues, and advancements in knowledge and technology. The curriculum must be guided by a college curriculum committee. The curriculum as a whole must be reviewed at least every seven (7) years. The majority of

Standard 11 - Outcomes Assessment

Outcomes of the veterinary medical degree program must be measured, analyzed, and considered to improve the program. New graduates must have the basic scientific knowledge, skills, and values to provide entry-level health care, independently, at the time of graduation. Student achievement must be included in outcome assessment. Processes must be in place to remediate students who do not demonstrate competence in one or more of the nine competencies.

What are your course outcomes?

Course Description:

A study of the etiology, clinical signs, diagnosis, treatment, and control of common medical diseases which affect horses.

Course Objectives:

- A. To gain an understanding of foundational aspects of equine internal medicine.
- B. To understand differences and similarities of equine medicine to small animal and ruminant patients.
- C. To understanding the clinical application of clinical internal medicine including diagnostic evaluation, clinical management, formulation of a prognosis and client communication that impact equine internal medicine patients.

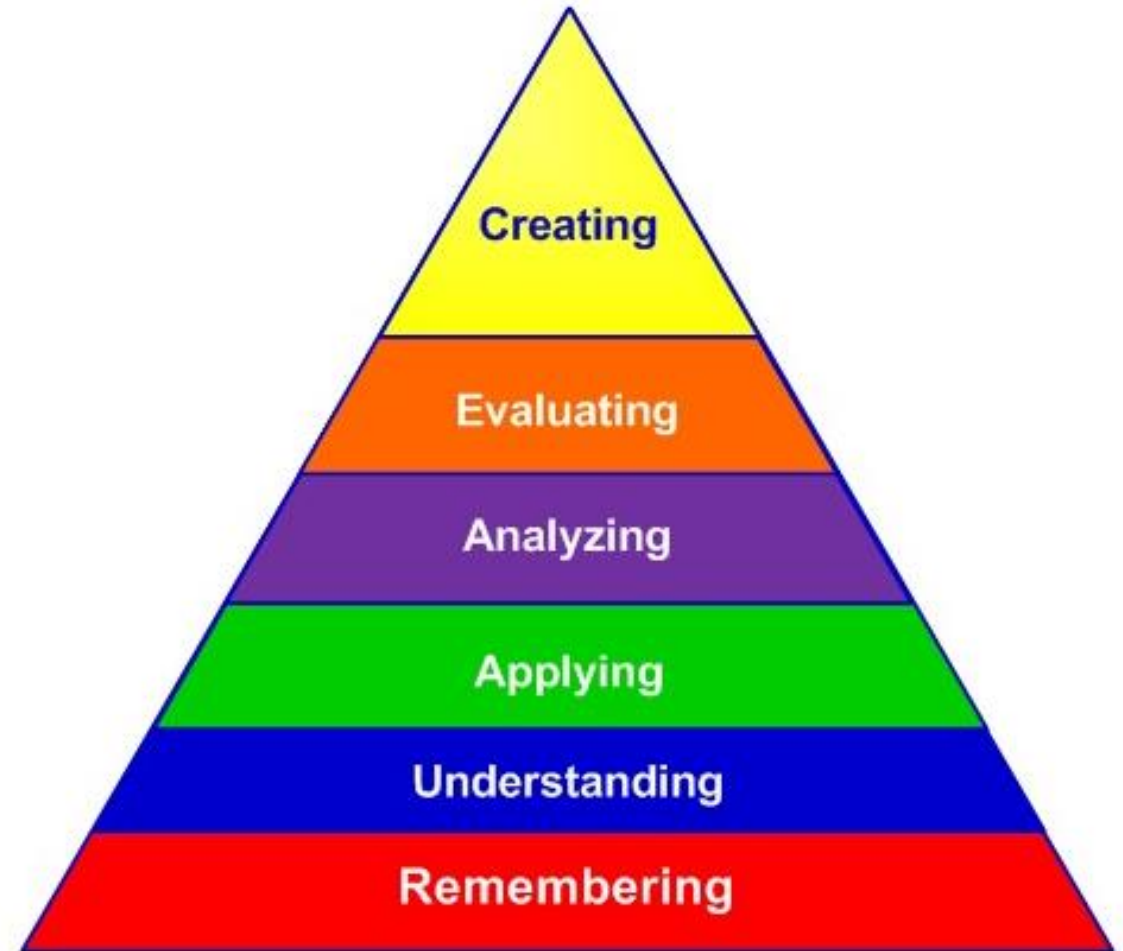
Lecture learning outcomes – core Equine Med

- Through the course of this lecture learners will be expected to know water distribution in the equine patient.
- Learners will be expected to calculate level of dehydration in equine patients based on clinical parameters that are provided.
- Learners will be expected to determine the most appropriate fluid therapy selection for an equine patient based on electrolyte status.
- Learners will be expected to calculate HCO_3^- and K^+ deficits in equine patients with electrolyte abnormalities.

Goals of the assessment

- **Course goals**
 - Describe the overall purpose of the course within the larger curriculum
 - Stated in general terms to encompass a domain of student performance
 - Provided in course syllabus
- **Learning objectives**
 - Break down goals into measurable and observable student performance that demonstrate competency
 - Ensure successful accomplishment of course goals
 - Ideally provided for each learning interaction

Blooms Taxonomy - Revised



Test / Examination Blueprint



The NAVLE®: From Survey to Blueprint

NAVLE

Author: Heather Case, DVM, MPH, DACVPM, CAE

The third part of our educational series explains how items are determined for inclusion on the North American Veterinary Licensing Examination (NAVLE®).

The NAVLE is based on a blueprint, which indicates the requirements for building the proper test that evaluates the knowledge, skills, and competencies needed by veterinarians entering private clinical practice. The blueprint is established by using the results of a practice analysis process (including verification by subject matter experts (SMEs) in veterinary medicine) and is periodically updated to reflect changes in the practice of veterinary medicine.

How is a blueprint update determined, and what does the process look like?

Since the International Council for Veterinary Assessment (ICVA) recently completed [a practice analysis study](#) and recoding project, let's take a closer look at how the process from survey to blueprint unfolds.

Step 1: Practice Analysis

A practice analysis begins with a survey that is distributed to thousands of veterinarians throughout the U.S. and Canada. The survey is used to obtain information about the species that veterinarians routinely see in practice and the diagnoses they are required to manage in their work.

Based on the results of the most recent practice analysis, an updated blueprint for NAVLE was created. This new blueprint was developed based on "competencies" (i.e., the ability to do something and understand why it should be done), rather than "activities" (i.e., a thing that a person does).

This shift from activities to competencies meant that current NAVLE items had to be recoded to reflect the new competency codes.



EXAM BLUEPRINT

The NAVLE is based on a blueprint created from a job analysis done by the ICVA.

The blueprint has two main parts:

1 Competencies

Competencies are the tasks that veterinarians perform in their daily work, and they are divided into four main categories. Each category is assigned a weight reflecting its importance in veterinary practice.

2 Species and Diagnoses

Species and Diagnoses pertains to the various animal species involved in veterinary medicine. Each species is assigned a weight based on how frequently they are encountered in veterinary practice in the U.S. and Canada.

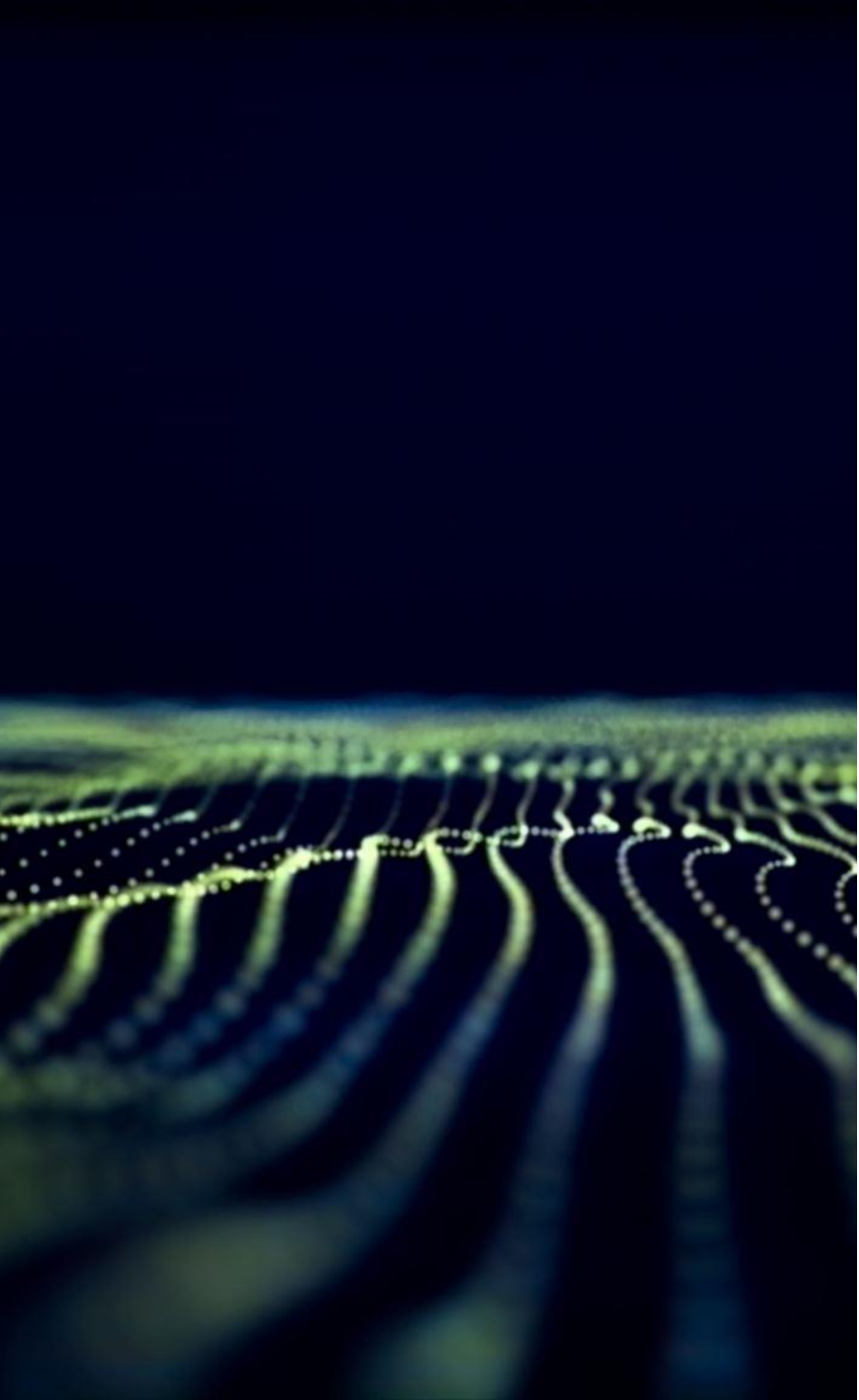
Competency Domains	Target Blueprint	Species	Target Blueprint
Clinical Practice	70%	Canine	25.6%
Data Gathering and Interpretation	35%	Feline	24.3%
Health Maintenance and Prevention	35%	Equine	14.7%
Preventive Medicine and Animal Welfare	15%	Bovine	13.3
Animal Welfare Issues and Concerns	6%	Porcine	5%
Environmental Health and Safety	5%	Other Small Mammals	3.3%
Veterinary Public Health	4%	Ovine/Caprine	3.3%
Communication	8%	Pet Bird	2.3%
Communication with Clients	5%	Poultry	2%
Communication with Veterinary and other Professionals	3%	Non-Species Specific	2%
Professionalism, Practice Management and Wellness	7%	Camelid/Cervidae	1.7%
Veterinary Practice Management	4%	Reptiles	1.5%
Professional Development and Life-Long Learning	3%	Aquatics	1%

Exam Blueprints: A Student-centric Approach to Assessment

📅 October 9, 2023 ✍️ Pradeep Malreddy, DVM



Exam blueprints are a teaching tool that is still shrouded in controversy. While some educators swear by their effectiveness, others argue that they limit creativity and could potentially facilitate cheating. The debate, however, often overlooks one crucial aspect: What do the students think? In anatomy courses, particularly those in professional programs, the stakes are high. Understanding the utility and impact of exam blueprints from a student perspective can be a game-changer. They can also help instructors to ensure that exams are fair and equitable.



Advantages of providing a test blueprint

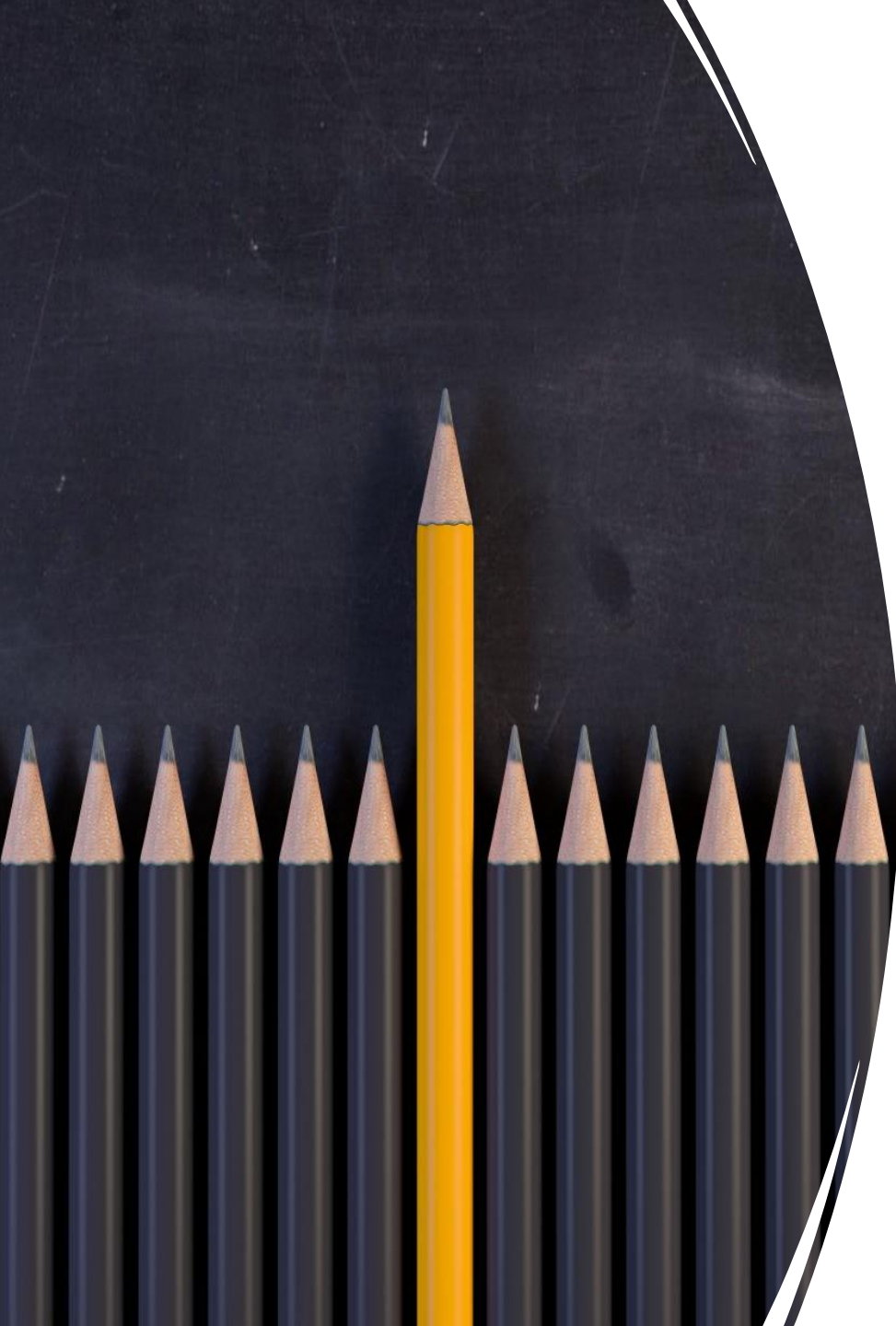
Consistency – assessment on same/similar criteria

Clarity – students understand the scope of examination

Improved learning outcomes – student understand the “big picture” of material

Fairness – avoid subjective grading practices

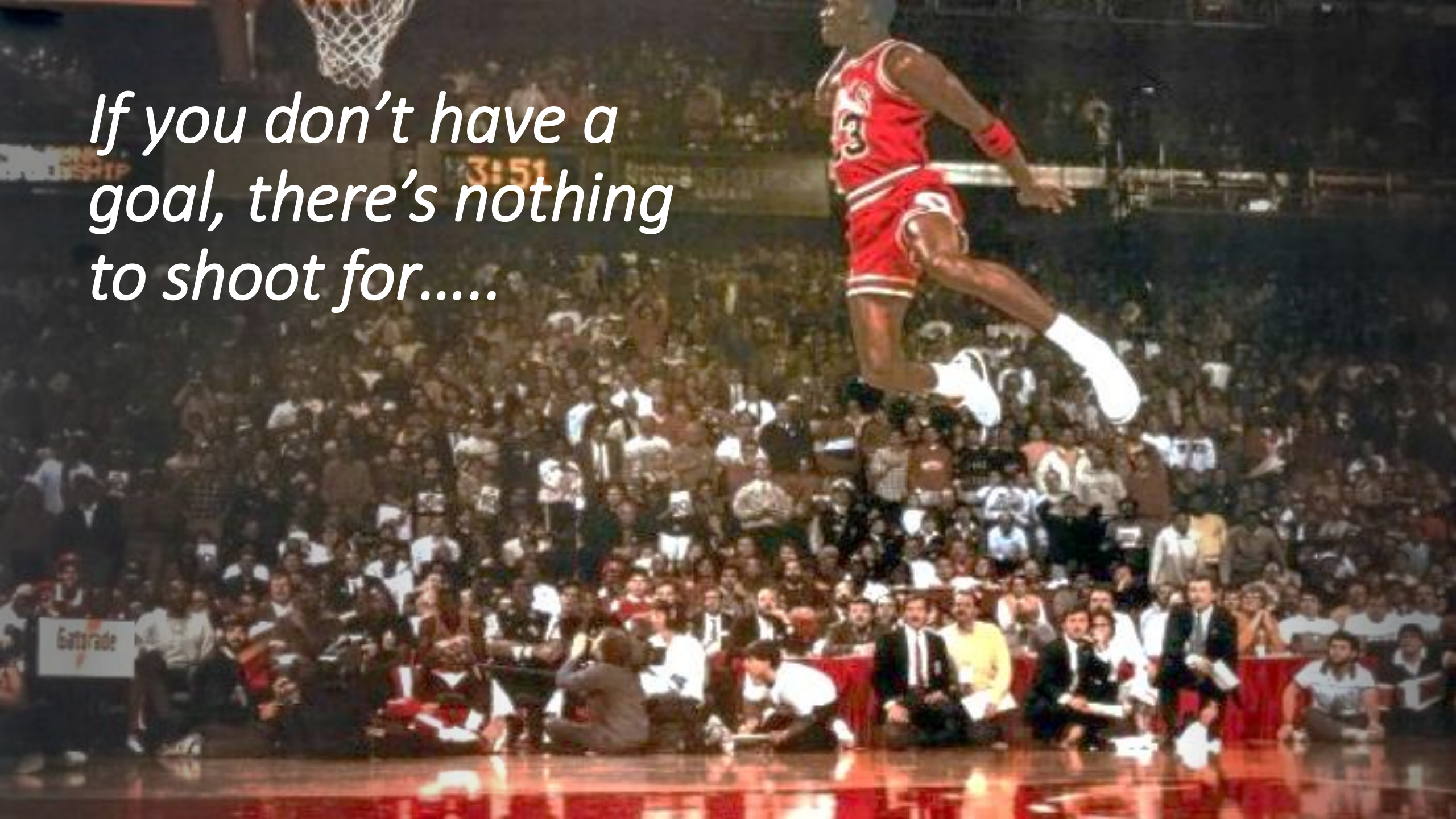
Better examination preparation – students can focus on key concepts

A decorative graphic on the left side of the slide. It features a row of pencils at the bottom, with one yellow pencil standing out in the center. Above the pencils is a large, dark, circular shape that resembles a blackboard or chalkboard.

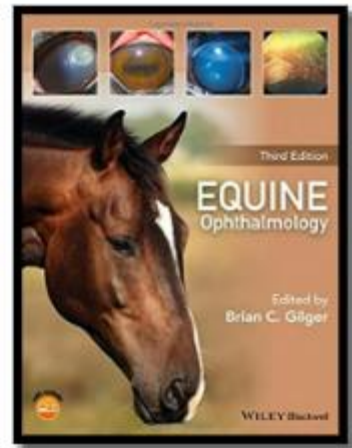
Disadvantages of test blueprint

-
- Limitation on creativity
 - Lack of unexpected questions
 - Decreased student motivation
 - Potential for shared material
 - Inability for faculty to change material to be examined

*If you don't have a
goal, there's nothing
to shoot for.....*



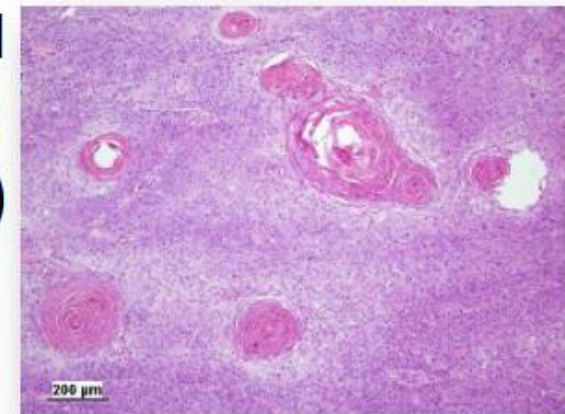
Learning Objectives



- Anatomy
 - Identify factors predisposing horses to SCC, corneal ulcers, corneal abscesses
 - Knowledge of landmarks for periocular nerve blocks
- Recognition and diagnosis
 - Recognize clinical signs associated with SCC, corneal ulcers, corneal abscesses
- Treatment modalities
 - Compare and contrast different treatment options including medical and surgical options
 - Evaluate the efficacy of medical treatments

SCC Treatment

- Therapeutic goal
- Surgical excision/debulking & adjunctive therapy
 - Cryotherapy
 - Carbon dioxide laser
 - Radiofrequency hyperthermia
 - Radiation therapy-strontium 90, cobalt 60, iridium 192
 - Chemotherapy – intralesional 5-fluorouracil, bleomycin or cisplatin (eyelid cases only)
 - Topical chemotherapy – 1% 5-fluorouracil
 - Immunotherapy (BCG) (eyelid cases only)
 - Photodynamic therapy (eyelid cases only)
 - Systemic piroxicam?



Attempts: 124 out of 124

+0.38

What is the **MOST** appropriate therapy for a squamous cell carcinoma of the lower eyelid in a horse?

Discrimination Index



Tattoo the eyelids		0 %	
Recommend the horse wear a Guardian mask	4 respondents	3 %	
Surgically debulk the tumor and cryotherapy	109 respondents	88 %	 ✓
Perform a wedge resection with 1 cm margins and submit for histopathology	11 respondents	9 %	

88% answered correctly

Bloom's Taxonomy

- **Teaching philosophy:** Benjamin Bloom, Taxonomy of Educational Objectives: the classification of Educational Goals. 1956.
- Classification of learning objectives into levels of complexity and specificity.
 - **Cognitive** – knowledge-based: objectives, assessment, activities.
 - **Affective** – emotion based.
 - **Sensory domains** – action based.
- Original publications in early to 1950s
- Revised taxonomy in 2001

Revised Bloom's Taxonomy

A group of cognitive psychologists, curriculum theorists and instructional researchers, and testing and assessment specialists published in 2001 a revision of Bloom's Taxonomy with the title *A Taxonomy for Teaching, Learning, and Assessment*. This title draws attention away from the somewhat static notion of "educational objectives" (in Bloom's original title) and points to a more dynamic conception of classification.

The authors of the revised taxonomy underscore this dynamism, using verbs and gerunds to label their categories and subcategories (rather than the nouns of the original taxonomy). These "action words" describe the cognitive processes by which thinkers encounter and work with knowledge.

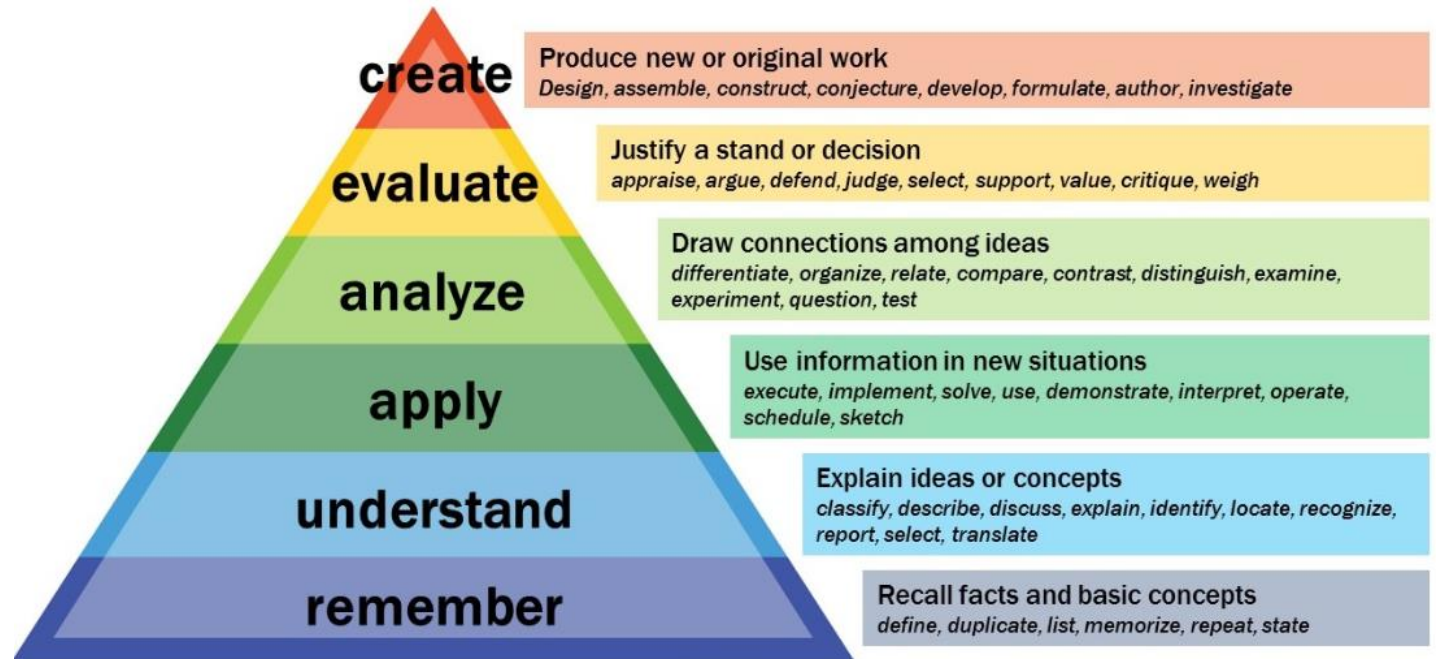
A statement of a learning objective contains a verb (an action) and an object (usually a noun).

- The **verb** generally refers to [actions associated with] the intended **cognitive process**.
- The **object** generally describes the **knowledge** students are expected to acquire or construct. (Anderson and Krathwohl, 2001, pp. 4–5)

The **cognitive process dimension** represents a continuum of increasing cognitive complexity—from remember to create. Anderson and Krathwohl identify 19 specific cognitive processes that further clarify the bounds of the six categories (Table 1).

Item
complexity

Bloom's Taxonomy



Examination Item Complexity

- Knowledge
- Comprehension
- Application
- Evaluate
- Synthesis / evaluation

- **Lower-level complexity**
 - Recall factual knowledge
 - Conceptual understanding
- **Higher level complexity**
 - Application of knowledge
 - Requires students to determine next step, say diagnostic test
 - Requires students to analyze
 - Evaluate procedural analysis

Recall or lower-level cognition

- **Factual recall:**
 - Which species does not have a gall bladder?
- **Conceptual recall:**
 - Which of the following conditions is an example of ventilation-perfusion mismatch?



Applied or higher-level cognition

- **Direct application**
- Which of the following diagnostic tests is indicated?
- **Analyze conceptual knowledge**
- Interpret bloodwork, radiograph or complex case scenario.
- **Evaluate case management**
- Case vignette, which steps were unnecessary?



Who determines “proper questions”

- Psychometrician assessment
- Practice the science of educational and psychological measurement = assessment or testing.
- Measure the validity, reliability and fairness of an exam program.
- Ensure that each part of the process conforms to testing industry standards.



UNIVERSITY OF
 CALGARY

Search UCalgary Profiles

UCALGARY PROFILES

People Directory
 Business Directory
 Emergency Info



Kenton Hecker

ON THIS PAGE

- Contact Information
- Background
- Research
- Courses
- Awards
- Publications

POSITIONS

Professor (Teaching & Research) - Veterinary Medical Education

Faculty of Veterinary Medicine, Fulltime

Adjunct Professor

Faculty of Arts, Department of Psychology

AFFILIATIONS

Professor, Joint Appointment

Cumming School of Medicine

Full Member

Hotchkiss Brain Institute

Full Member

Mathison Centre for Mental Health Research and Education

Full Member

O'Brien Institute for Public Health

Dr. Hecker has built a research program of basic and applied health professional education research focusing in four areas: 1) curriculum/program design and evaluation; 2) assessment of competencies; 3) educational measurement and assessment issues within institutional and non-institutional settings; and 4) the application of neuro-imaging methods (functional magnetic resonance imaging [fMRI] and electroencephalography [EEG]) to assess learning, reasoning and decision making.

Pioneering Excellence in Educational Research and Design

UCLA CRESST is a renowned center with decades of groundbreaking work in the field of educational research. Our work on assessment, evaluation, educational technology, and methodology shaped current knowledge and moves the field forward.



National Center for Research on Evaluation, Standards and Student Testing (CRESST)

- Cognitive complexity
- Content quality
- Meaningfulness
- Language appropriateness
- Transfer and generalizability
- Fairness and reliability

Quiz Summary

Section Filter ▾

 Student Analysis

 Item Analysis

Ⓜ Average Score

86%

📈 High Score

100%

📉 Low Score

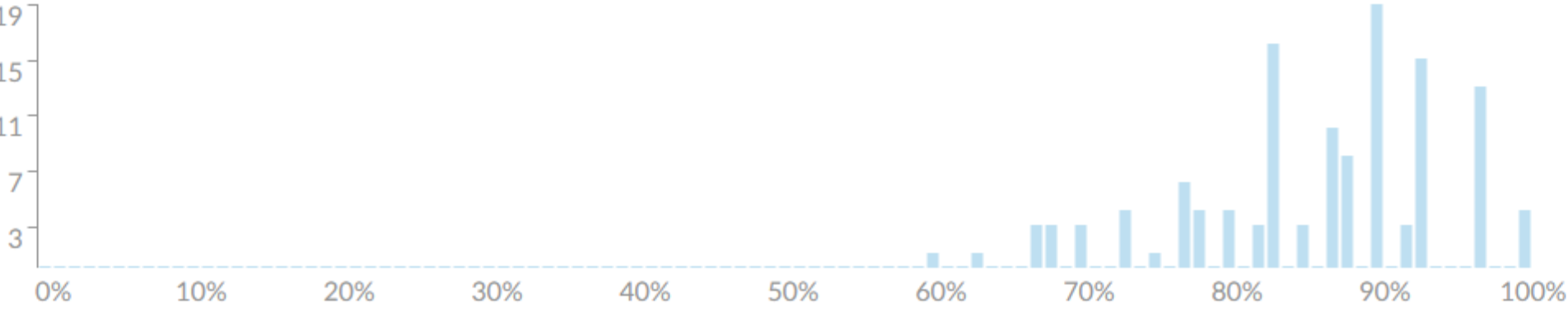
60%

Ⓢ Standard Deviation

10.43

🕒 Average Time

40:00



A Taxonomy of Multiple-Choice Item-Writing Rules

Thomas M. Haladyna
*Arizona State University
West Campus*

Steven M. Downing
National Board of Medical Examiners

A taxonomy of 43 multiple-choice item-writing rules is presented and discussed. The taxonomy derives from an analysis of 46 authoritative textbooks and other sources in the educational measurement literature. The analysis also leads to a "validity by consensus" for each rule. The taxonomy is viewed as a complete and authoritative set of guidelines for writing multiple-choice items.

Is This a Trick Question?

A Short Guide
to Writing Effective
Test Questions

Designed & Developed by:
Ben Clay
Kansas Curriculum Center

Formatting & Text Processing by:
Esperanza Root

This publication was developed by the
Kansas Curriculum Center with funds
provided by the Kansas State
Department of Education.
First printing:
October, 2001

Is This a Trick Question?

A Short Guide
to Writing Effective
Test Questions



COLLEGE OF VETERINARY MEDICINE

Hospitals & Services >

Current Students >

About ▾ Education ▾ Research ▾ Centers & Programs ▾ Departments ▾ Alumni ▾ Donors ▾ News & Events ▾ Give

Home > Education > Educational Support > Teaching Skills > Assessment > Writing Test Questions

Writing Test Questions

DVM Program

MS & PhD Programs

Dual Degree Programs

DVM/MPH 

Continuing Education

Educational Support

▼ CVM Educational Support Group

Canvas

Clinical Teaching

Course Coordination

DVM Curriculum

Educational Research

Evaluation of Teaching

Faculty Development
Opportunities

See the article below about choosing the right kind of assessment tool including a good checklist for instructors when looking at their assessments. The other article is an analysis of current research about how we write test questions and what are best practices.

Articles

 [Choosing the right assessment tools](#)

 [A taxonomy of multiple-choice item writing rules](#)

There are two main types of test questions, defined response questions and constructed response questions. Defined response questions include true-false, matching, and multiple-choice questions. Constructed response questions include short-answer, essay, and oral examinations.

True-false questions seem easy to write but since most statements are not absolutely true or absolutely false, it can be difficult to write completely unambiguous true-false examinations. True-false questions tend to emphasize recall of isolated facts and often cannot be used to test higher level learning. The same is true for matching questions. Multiple-choice questions are the most common testing format and, if well constructed, can be used to test higher order learning. Some students will excel at multiple-choice examinations not because they have a great depth of knowledge but because they are good at finding clues in how assessments are written. Instructors always should strive to ensure assessments are permitting students to demonstrate the level of learning described in the course learning objectives. Here is a brief review of things to think about when writing multiple-choice questions.

Multiple choice questions (MCQs) are written with a stem and several possible answers, or foils. The following are basic rules of MCQ construction:

Multiple choice questions written as stem and several possible answers = distractors or foils:

- Put the central idea in the stem, avoid repeating in answers / distractors.
- Avoid negative construction, “which of the following is not correct”.
- Use proper grammar and ensure the correct foil is not obvious because of grammar clues.
- Do not use terms always or never in foils.
- Place foils in logical or numerical order.
- Make sure foils are parallel with respect to grammatical structure, length and complexity.

Composition of the stem

Stem should be a complete sentence.
Stand-alone problem.

Students should be able to determine the
response / outcome based on the stem.

Avoid awkward statements.

Eliminate extraneous or misleading
information.

Constructing questions

- Use “none of the above” with caution; do not use “all of the above” as a foil.
- Three distractors / foils are the ideal number.
- To prepare reasonable distractors consider common student misconceptions.
- Do not worry a great deal about distribution of responses, the evidence does not support that students determine patterns for answers.
- Avoid or use humor sparingly.

Items more difficult than intended

- Awkward stem structure. Finish the sentence, fill in the blank, grammatically flawed.
- Extraneous or misleading information in the stem.
- Negative stem, not true, true except, incorrect.
- Response options are a series of true/false statements.
- Use of none of above.
- Complex or K-type items. Example = A and C.
- Vague or generalizing terms. Sometimes, frequently, often, occasionally, typically or potentially.
- Unfocused question. Distractors are unrelated or distantly related to a single objective.



Items are easier than intended



- Longest option is correct answer.
- Grammatical clues or inconsistency between stem and distractors.
- Implausible distractors.
- Mutually exclusive distractors.
- Use of absolute terms (always, never, only, all).
- Use of all of the above.

Avoid negatively worded multiple-choice stems

Although simpler to construct, positively worded questions more effectively assess the students ability to learning objectives.

A stem and response option leading to a double negative is particularly challenging for students to answer correctly.

Education in medical health fields does lend itself to having students understand inappropriate = contraindicated treatments. In this instance **CONTRAINDICATED** should be in bold and underlined for clarity.

Multiple choice variants

- Series of true / false statements.
- Which of the following statements is correct / incorrect?
- Students are responding to a series of true false questions with all or none grading rather than assessment of one multiple choice item.
- Selected by instructors to cover a larger amount of material on examination.
- Disadvantages students that have knowledge of the tested content.

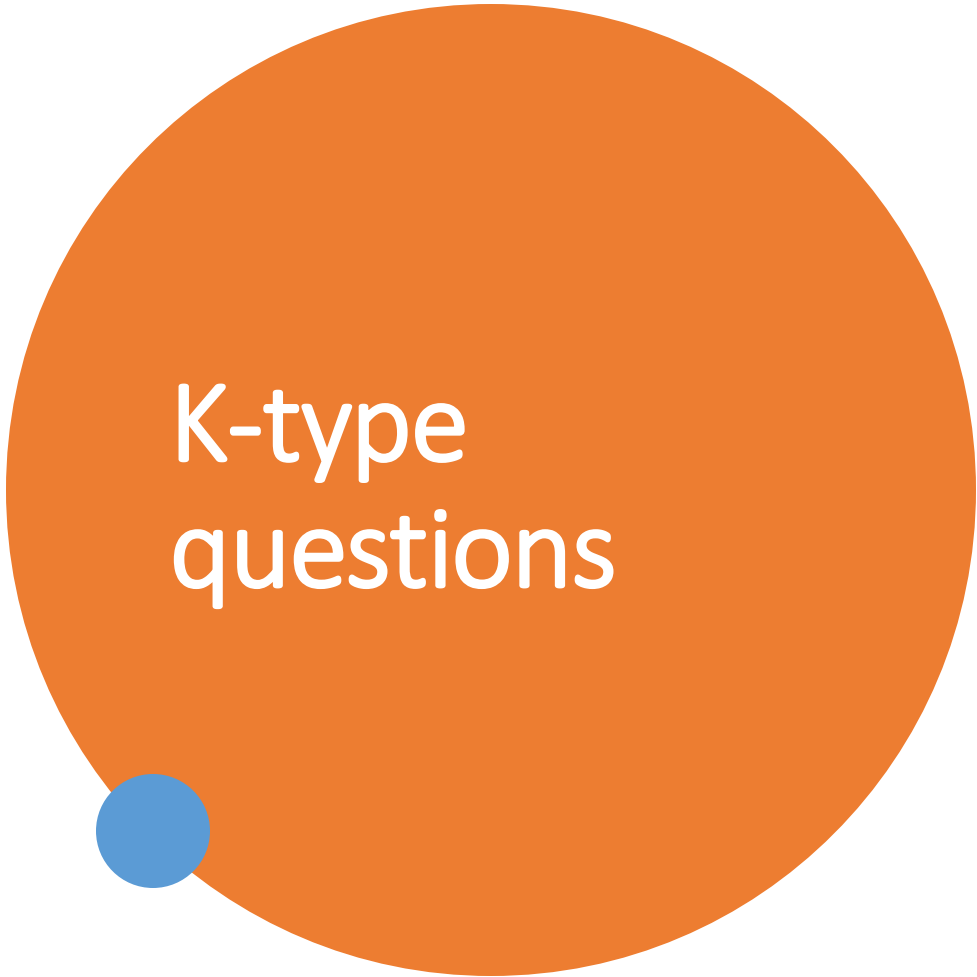
“None of the above”

Does not require
knowledge of
correct answer.

Often used as a
filler.

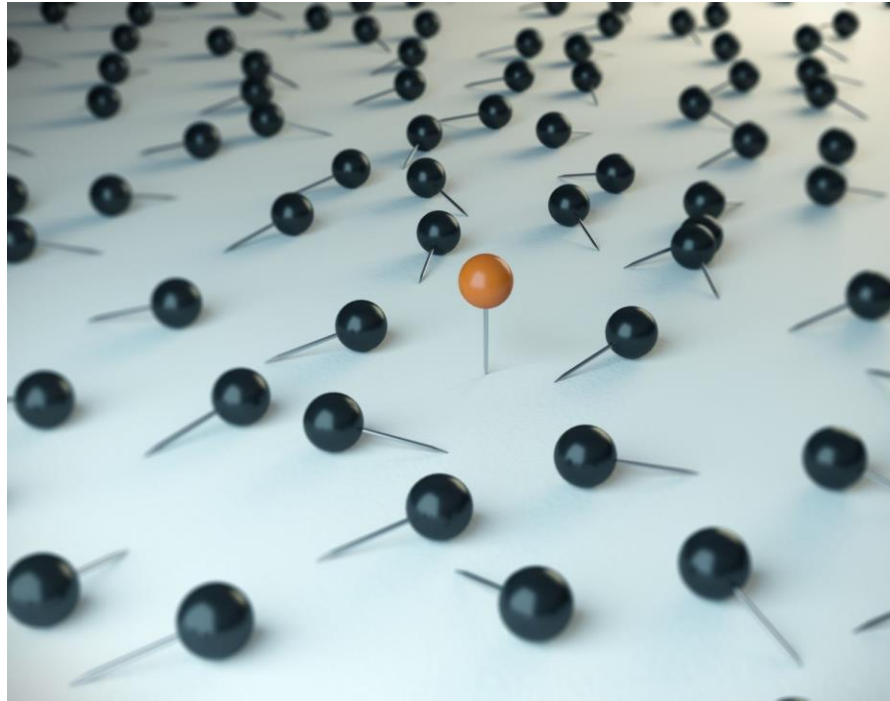
Reduced
discrimination and
test reliability.

Suggested to avoid
this distractor.



K-type questions

- Complex items.
- Requires a student to select combinations of individual response options.
- Generally a series of true / false options.
- Often contain grammatical clues that serve as an advantage for experienced test takers.
- Evidence demonstrates that K-type questions perform poorly in reliability and discrimination when compared with single correct answer and true / false formats.



Avoid vague or generalizing terms

- Frequently, often, sometimes, occasionally can have different interpretations for different audiences.
- This type of question often has an unfocused stem with a broad open ended question that does not pose a specific problem and is followed by a series of unrelated response options.
- Examiner can test a broad range of material but does not provide an assessment of a specific learning objective.
- This form of question has a detrimental impact on student and item performance.

True / false questions

Advantages

- Straightforward to prepare.
- Students can respond to a large number compared with multiple choice items.

Disadvantages

- Students can guess correctly 50% without knowledge of the content.
- Licensing examination boards do not allow T/F questions.
- False items are more discriminatory.
- Students are more likely to guess true.

View Discrimination Index

Attempts: 3 out of 4

How much time should you spend on your online course each week in order to understand the concepts?

+0.38

Discrimination Index ?

30 minutes	1 respondents	25 %	✓
At least 3 hours		0 %	
However long it takes		0 %	
1-2 hours	2 respondents	50 %	
No Answer	1 respondents	25 %	



Attempts: 113 out of 113

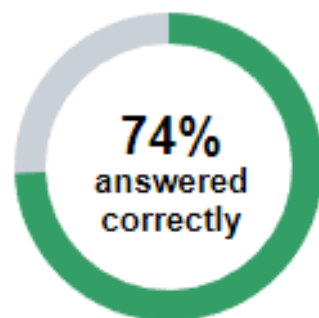
+0.40

Discrimination Index



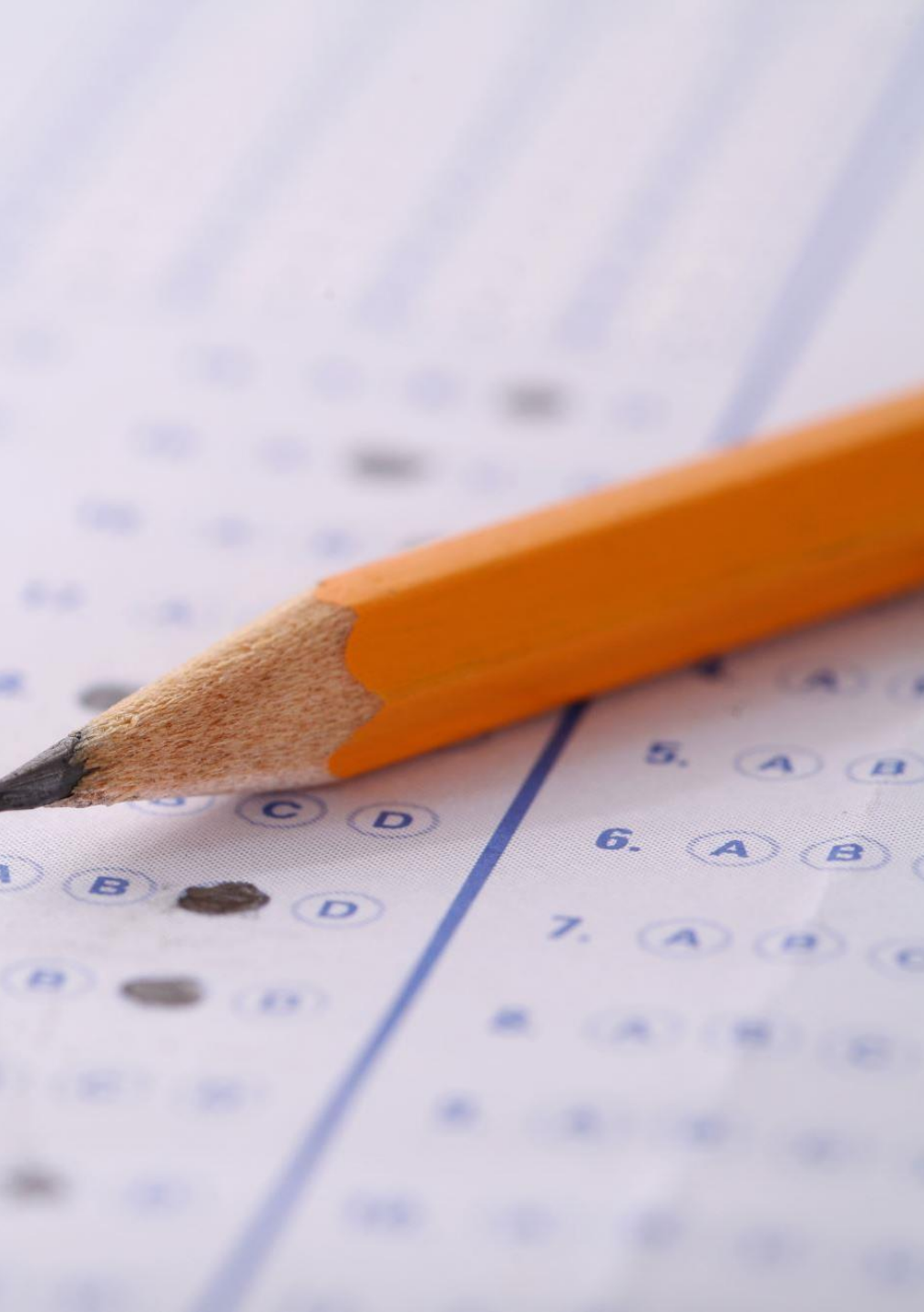
Which of the following actions is likely to speed the onset of anesthetic induction with an inhalation anesthetic?

Using a lower fresh gas flow rate (O2 flowmeter) on a Circle system so there is less dilution of inhalant	12 respondents	11 %	
Switching to an inhalation agent with higher blood:gas partition coefficient value	8 respondents	7 %	
Administering a positive inotrope such as dobutamine to the animal to increase its cardiac output	9 respondents	8 %	
Placing the animal on IPPV using a mechanical ventilator to increase ventilation	84 respondents	74 %	✓



Discrimination Index

- Three test taker groups
 - Top 27%
 - Middle 46%
 - Bottom 27%
 - Correct in bottom subtracted from top, remainder is divided by total
- Lower discrimination scores are scored **+0.24** or lower
- Good scores are **+0.25** or higher
- An ideal discrimination index shows students who scored higher on the quiz get the question correct.
- Students who scored lower on the quiz getting the quiz question wrong.
- Students in the middle on both sides.
- Discrimination of zero shows all students got the question right or wrong.



- **Discrimination index:**
- Measure of how well a question can tell the difference (discriminate) between students who do well on an exam and those who do not
- Test takers into three groups based on their score on the whole quiz and displays those groups by who answered the question correctly.
- Goal is to have students who knew material well perform well and those that do not (or are guessing) not correctly answer the question.

3. If managing a foal confirmed with glycogen branching enzyme deficiency (GBED), what is the likelihood that the dam is a carrier for this condition?

- a. 25%
- b. 50%
- c. 75%
- d. 100%

Attempts: 115 out of 115

Q3

a	8 respondents	7 %	
b	18 respondents	16 %	
c	4 respondents	3 %	
d	85 respondents	74 %	✓

+0.27

Discrimination Index



8. Choose the correct statement regarding ERU.

- a. ERU causes signs of ocular pain in the active stage but rarely results in blindness.
- b. The causative agent for ERU is equine herpesvirus type 1.
- c. The *posterior* clinical classification produces the most obvious clinical signs of ERU.
- d. Appaloosa and draft breeds are at an increased risk of ERU development.

Attempts: 115 out of 115

Q8

a	4 respondents	3 %	
b		0 %	
c	3 respondents	3 %	
d	108 respondents	94 %	✓

+0.37

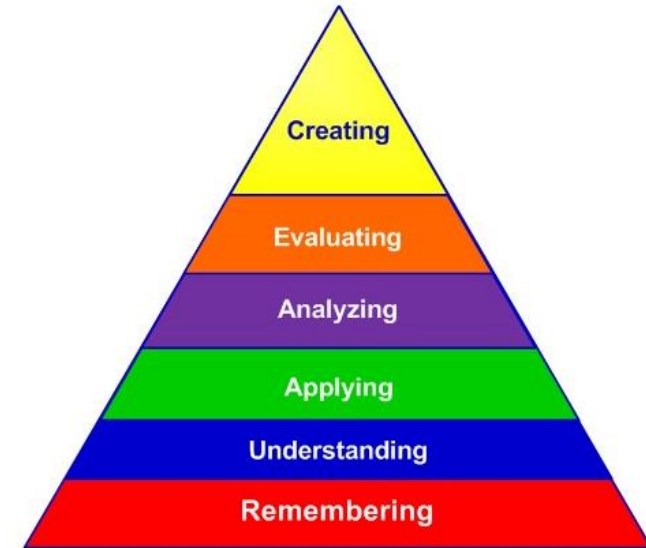
Discrimination Index



Question preparation 2026

- Assistive or non-generative AI
- Prompt preparation
- Enter notes or prepared material
- 20% of the content recall/recite
- 80% higher level – interpret, create, analyze
- Prepare prompt to assess course material based on outcomes
- **Carefully** proof material for accuracy and consistency with delivered content

Blooms Taxonomy - Revised



Prompt:

- Third year veterinary students, multiple choice
 - Recall - equine respiratory examination lecture
-
- **Extended head and neck posture with loud inspiratory noise most strongly suggests**
 - A. Lower airway obstruction
 - B. Restrictive lung disease
 - C. Pleural effusion
 - D. **Upper airway obstruction**

Prompt - creation and application

During examination of a horse with respiratory distress, you observe extended head and neck posture, inspiratory noise, and no improvement during rebreathing. However, when you perform a temporary tracheotomy, the horse's breathing improves immediately.

Which finding best explains why the tracheotomy resolved the clinical signs?

A. The horse had intrapulmonary restrictive disease that improved with larger tidal volumes

B. The obstruction occurred in the extrathoracic airway, bypassed by the tracheotomy

C. The horse had pleural effusion causing diaphragmatic restriction

D. The horse was experiencing dynamic intrathoracic airway collapse worsened during expiration

Prompt - application of knowledge

A 12-year-old gelding presents with increased expiratory effort, a pronounced abdominal press, and wheezes that are loudest at the end of expiration. On auscultation, you note prolonged expiration and an increased functional residual capacity during rebreathing.

Based on these findings, which pathophysiologic process is MOST consistent with this horse's presentation?

- A. Upper airway dynamic collapse due to laryngeal hemiplegia
- B. Intrathoracic small airway obstruction causing air trapping**
- C. Restrictive pulmonary disease reducing lung compliance
- D. Extrathoracic airway narrowing causing inspiratory stridor

Measures for Quality Assurance of Electronic Examinations in a Veterinary Medical Curriculum

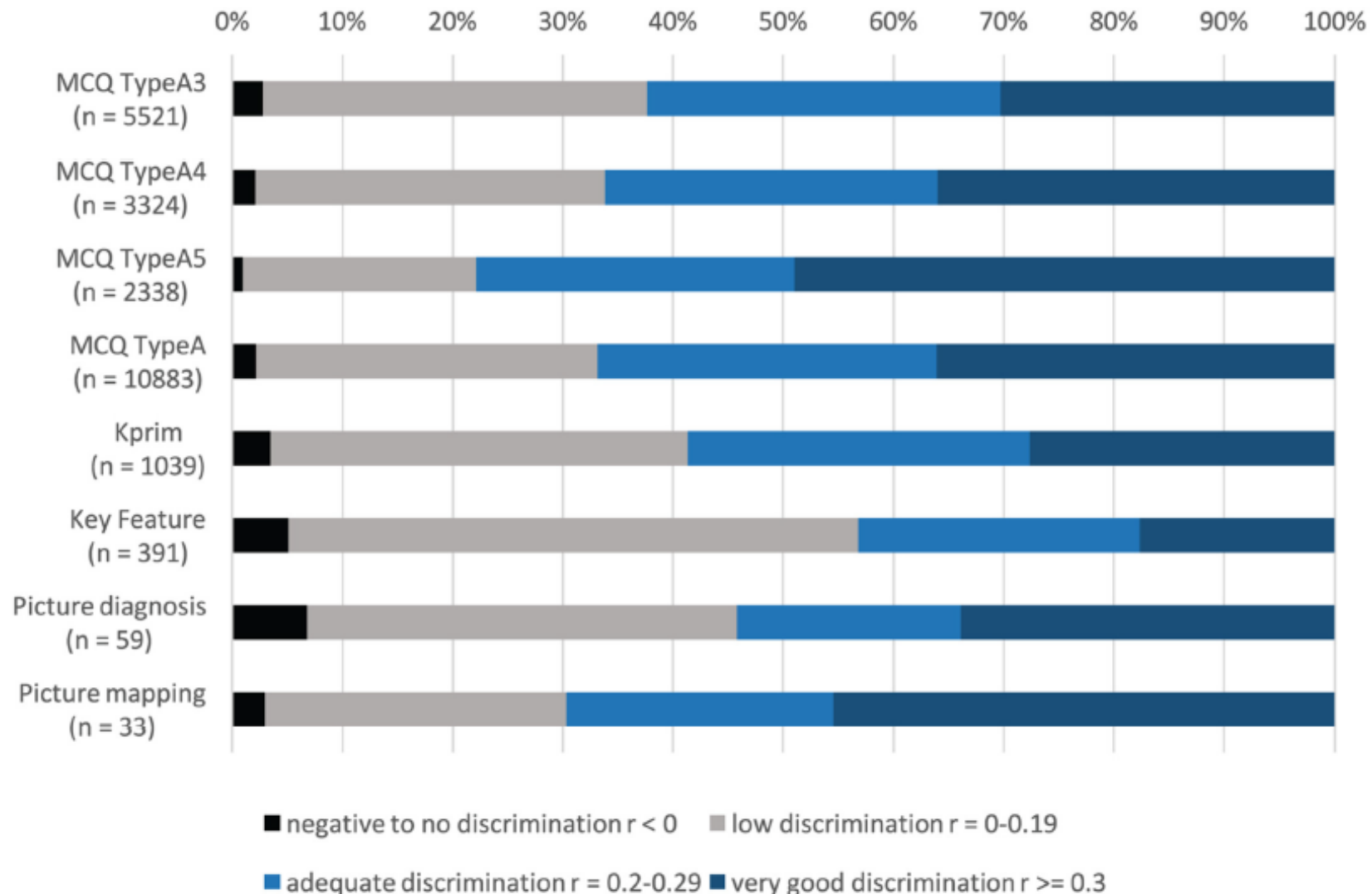
Robin Richter ■ Andrea Tipold ■ Elisabeth Schaper

ABSTRACT

Since 2008, electronic examinations have been conducted at the University of Veterinary Medicine Hannover, Germany which are analyzed extensively in the current study. The aim is to assess the quality of examinations, the status quo of the electronic examination system and the implementation of recommendations regarding the conduct of exams at the TiHo. Based on the results suitable indicators for the evaluation of examinations and items as well as adequate quality assurance measures and item formats are to be identified. For this purpose, 294 electronic examinations carried out from 2008 to 2022 of the veterinary medicine course with an average of 248 participants each were evaluated with regard to the quality criteria reliability, difficulty index, and discrimination index. The main finding was that the number of items and the proportion of reused questions were identified as factors through which the quality of the examinations can be increased with simple adjustments. A higher number of items led to better reliability, whereby the required minimum reliability in examinations of 0.8 was reliably achieved from an item number of 98 questions. The proportion of reused questions should be kept low, as these had a negative influence on the characteristic values. Measures accompanying examinations, such as training of question authors and a pre- and post-review process, should also ensure the quality of examinations. For the post-review process, the distribution of examination results, reliability, item and distractor analysis are adequate indicators for evaluating examinations.

Key words: e-assessment, veterinary education, examinations, item formats, Cronbach's α , discrimination index, multiple choice questions

Discrimination index separated by item format





The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value

Bonnie R. Rush*, David C. Rankin and Brad J. White

Abstract

Background: Failure to adhere to standard item-writing guidelines may render examination questions easier or more difficult than intended. Item complexity describes the cognitive skill level required to obtain a correct answer. Higher cognitive examination items promote critical thinking and are recommended to prepare students for clinical training. This study evaluated faculty-authored examinations to determine the impact of item-writing flaws and item complexity on the difficulty and discrimination value of examination items used to assess third year veterinary students.

Methods: The impact of item-writing flaws and item complexity (cognitive level I-V) on examination item difficulty and discrimination value was evaluated on 1925 examination items prepared by clinical faculty for third year veterinary students.

Results: The mean ($\pm$ SE) percent correct ($83.3\% \pm 17.5$) was consistent with target values in professional education, and the mean discrimination index (0.18 ± 0.17) was slightly lower than recommended (0.20). More than one item-writing flaw was identified in 37.3 % of questions. The most common item-writing flaws were awkward stem structure, implausible distractors, longest response is correct, and responses are series of true-false statements. Higher cognitive skills (complexity level III-IV) were required to correctly answer 38.4 % of examination items. As item complexity increased, item difficulty and discrimination values increased. The probability of writing discriminating, difficult examination items decreased when implausible distractors and all of the above were used, and increased if the distractors were comprised of a series of true/false statements. Items with four distractors were not more difficult or discriminating than items with three distractors.

Conclusion: Preparation of examination questions targeting higher cognitive levels will increase the likelihood of constructing discriminating items. Use of implausible distractors to complete a five-option multiple choice question does not strengthen the discrimination value.

Keywords: Multiple choice examinations, True-false examinations, Pre-clinical education, Item-writing flaws, Veterinary education

Background

- The goal of examination item preparation is to develop testing methods that **do not confuse students** and yield scores that accurately reflect the extent to which students have obtained satisfactory working knowledge of the content.
- Well-constructed multiple-choice examinations represent a versatile assessment tool with the potential to assess students for sufficient working knowledge of the tested content.
- Item writing flaws may render the examination easier or more difficult than intended.
- Clues or challenges should be avoided.

Methods

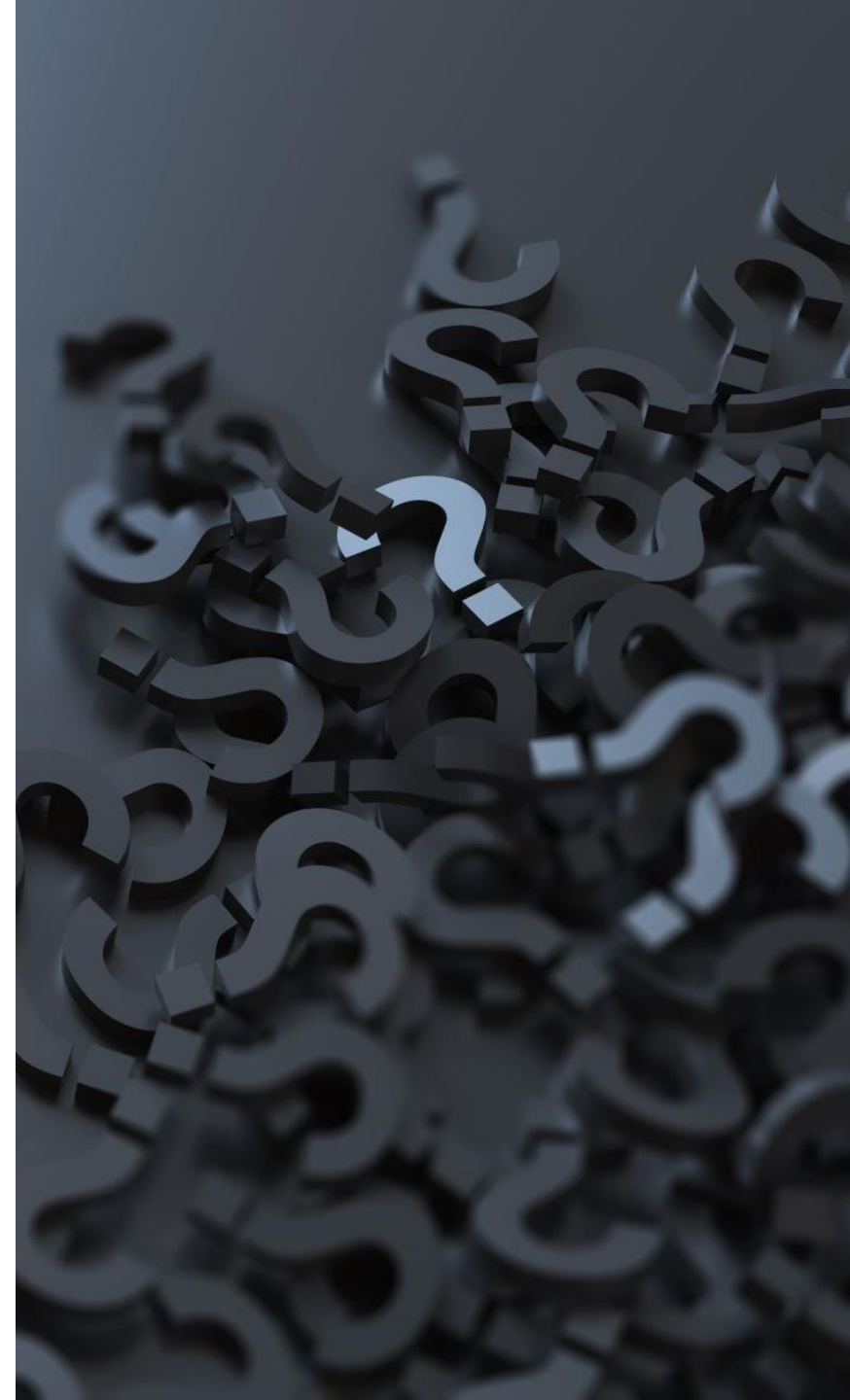
- 112 students: 85 female, 27 male (avg. age = 25.9 yrs)
- All examination questions in third year core courses
- Average of 7.2 years post-secondary education
- CVM faculty members composed questions
- Two examination readers examined and scored all items independently
 - Identified item writing flaws
 - Item complexity (I-V)
- Small animal medicine = 8 credits
- Small animal surgery = 5 credits
- Large animal medicine = 7 credits
- Large animal surgery = 4 credits
- Nutrition = 2 credits
- Reproduction = 3 credits
- Zoological medicine = 2 credits
- Ethics / jurisprudence = 1 credit
- Clinical pharmacology = 2 credits
- Practice management = 1 credit

Findings

- 1925 examination items
- Authored by 50 faculty members
- 46 examinations in 16 third year courses:
 - 12 core and 4 elective = 39 credits
- 1689 questions were multiple choice
- 236 were true / false

Results

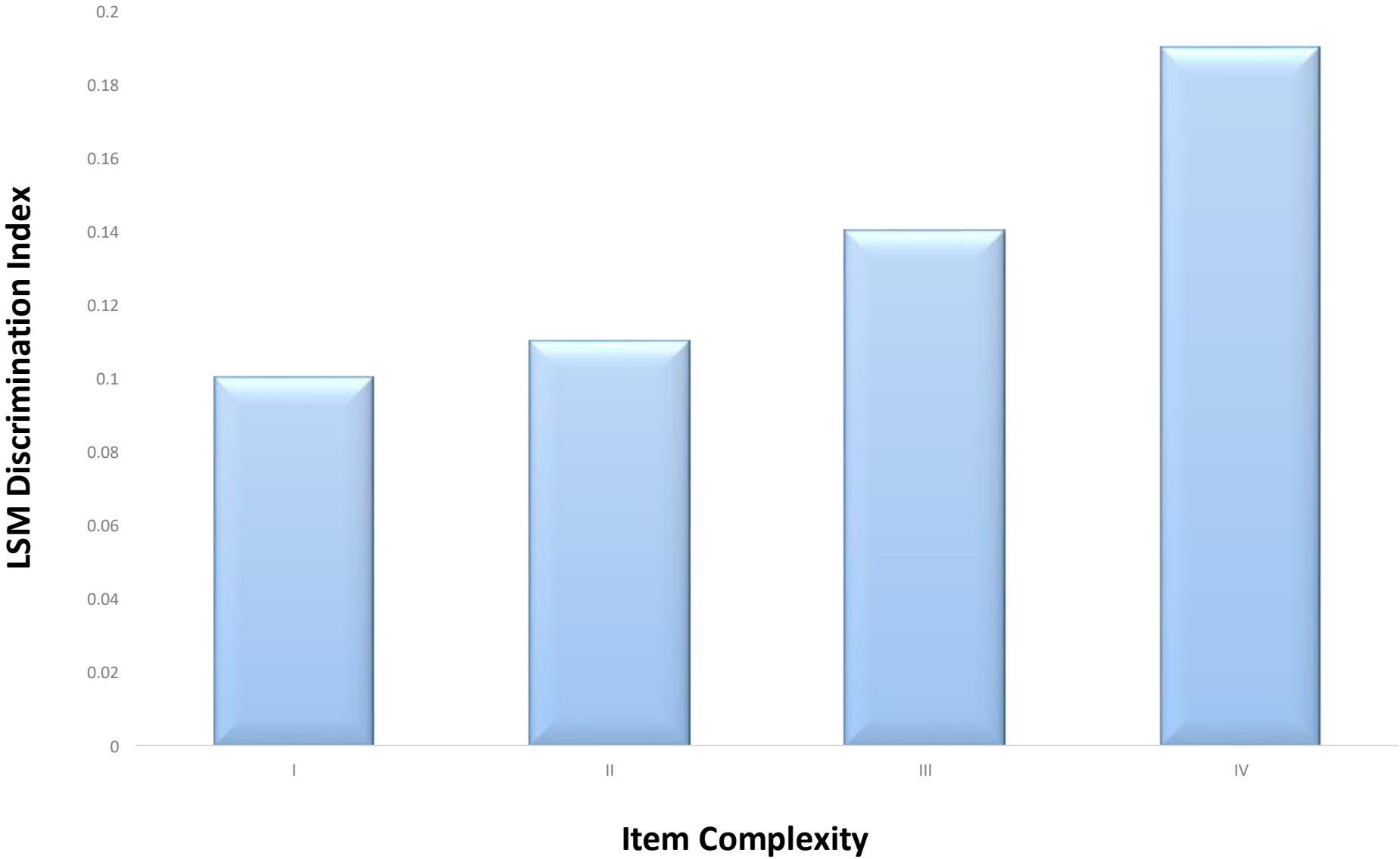
- Approximately **28.8%** of examination items were identified as free of item-writing flaws
- **33.9%** had one flaw
- **37.3%** had more than one flaw
- Author bias was not evident with regard to placement of correct answers: A – 25.1%, B 25.8%, C 23.9%, D 25.1%.
- True / false: 111 were true and 126 false.
- Lower recall: 61.6%
- Higher recall: 31.4%
- Highest recall: 6.9%



Discriminating questions

- Approximately 30.2% (573) exam items were classified as difficult, discriminating ($DI > 0.20$)
- As item complexity increased the likelihood of discriminating question increased
- The probability of writing difficult, discriminating questions decreased when
 - “all of the above” was a distractor
 - A series of “true / false” statements were provided

Association of item complexity and item discrimination



Common flaws with item preparation

- **This study:**
- Awkward stem structure
- Implausible distractors
- Longest response = correct
- **Other studies:**
- Unfocused questions
- Negatively worded stem
- Implausible distractors
- All of the above, none of the above and K-type distractors



Recommendations

- Higher cognitive level questions will be more discriminating
 - Identify student who know information from those who do not or guess
- Avoid “padding” examination questions with implausible distractors merely to ensure same number of distractors through examination
- Avoidance of true / false responses as this item format may disadvantage students

Summary

- Use course goals and learning outcomes to help students identify primary goals of course material.
- When preparing assessment methods, use objectives to establish the test blueprint.
- Write well developed knowledge/application-based response questions.
 - Traditional AI prompts may help with question writing
- Questions with higher degree of complexity that include clinical vignettes will provide better assessment of student knowledge.
 - Goal is to have 60-80% questions higher level complexity.
- Review test statistics to determine how questions perform.
 - Those with particularly low discrimination (<0.10) index should be reviewed, revised or removed.

