

ent
self
ble.
are
as
ng.
ion
rell
ed
be
in
on
to
a
w.
ion
of
nt.
nas
ont
ach
HE
tes,
to
ng
be
in
ner
of
nas
e a
en
the
be

**BEYOND THE *MULTIPLE-CHOICE* v. *ESSAY*
QUESTIONS CONTROVERSY:
COMBINING THE BEST OF BOTH WORLDS***

By ERIK DRIESSEN, CEES VAN DER VLEUTEN and
HENK VAN BERKEL†

1. Introduction

ACHIEVEMENT TESTS play an important role in (legal) education. Through tests we check whether our students have mastered the objectives of our educational programme. In addition, qualifying tests have a very strong influence on the study behaviour of students.

It is therefore extremely important that test content and format are consistent with the educational objectives.¹

Terence J. Crooks reviewed results from 14 specific fields of research on the impact of classroom evaluation practices on students.² He concluded that achievement tests have a strong direct and indirect influence on learning strategies, motivation and achievement of students, both positive and negative. On the issue of the test format Crooks concluded:

. . . there is no strong evidence from this research to support widespread adoption of any one item format or style of task. Instead, the basis for selecting item formats should be their suitability for testing the skills and content that are to be evaluated.³

In reality, most teachers and administrators continue to debate the format of test questions. Beliefs and misconceptions with regard to multiple-choice questions and essay questions dominate this debate. In legal education, there seems to be substantial support for essay questions. In a survey of law lecturers in England and Wales Tribe and Tribe found that, in spite of large differences in teaching objectives and criteria for assessment, the types of assessment technique varied remarkably little.⁴ That is, 91.6% of law lecturers used essay questions and written problems. Only 7% used objective/multiple-choice questions. In general, there seems to be quite a distrust of such objective question forms as the

* We would like to thank Jan Smits for his comment on the earlier draft of this article and Louise Rayar for revising the English of this article.

† Faculty of Law, University of Maastricht.

1 N. Frederiksen, "The Real Test Bias: Influences of Testing on Teaching and Learning", 39 *American Psychologist*, 193-202 (1984).

2 T. J. Crooks, "Impact of Classroom Evaluation Practices on Students", 58 *Review of Educational Research*, 438-481 (1988).

3 *Ibid.*, at 448.

4 D. Tribe and A. Tribe, "Assessing Law Students: Lecturers' Attitudes and Practices", Vol. XIII, *Assessment and Evaluation in Higher Education*, 195-211 (1988).

multiple-choice question format.⁵ The evidence in the extensive research on the effects of question format is more sobering: format does not matter as much as we tend to believe. The same attributes can be measured with the aid of either question format.⁶ An exception should be made for tests assessing writing skills, creativity or evaluation skills. These skills cannot be assessed with a multiple-choice question format.

The purpose of this article is to examine the feasibility of a test that combines the two question formats in higher legal education. We will seek an answer to the following questions:

Is it possible to develop a test with both open and closed questions which can assess knowledge of the law, insight and skills at a higher ability level? Is such a test more efficient than an all-essay test? How does it affect the study behaviour of students? And will students and faculty accept an assessment form containing both open-ended and closed questions?

2. Assessment System at the Maastricht University Faculty of Law

In the Netherlands, the first year of graduate law studies has two functions. First, to acquaint students with basic Dutch Law and, second, to select those students who are competent to study law at graduate level. In the first year, the Faculty of Law of Maastricht University offers four eight-week courses. The first course serves as an introduction to Dutch Law and the subsequent three courses deal with the six main areas of Dutch Law. The courses have a problem-based learning format, which means that tutors guide small groups of students who study on the basis of problems offered in the study materials. Different domains are more or less integrated within a single course.⁷

Parallel to the courses there are weekly skills-training sessions.

Up to 1995, each course was evaluated by an achievement test: an all-essay test consisting of four to six questions often on the basis of a case vignette, some with sub-questions. Students also had to complete a legal writing test at the end of their first year.

5 Many of the arguments against multiple-choice questions are expressed in an essay by E. Rule: "The Use of Objective Testing for Law School Admissions", 24 *The Law Teacher*, 37-47 (1990). She advocates research on objective tests in legal education. Our article is an example of such a study, except that we also include the open question format.

6 There is a long research tradition by both teachers and test specialists on the question as to whether essay tests measure something quite different than tests with multiple-choice questions. Two empirical studies of good quality about this topic with knowledge and cognitive problem-solving tests are: G. R. Norman *et al.*, "Factors underlying performance on written tests of knowledge," 21 *Medical Educ.*, 297-304 (1987); R. E. Bennett, D. A. Rock and M. Wang, "Equivalence of free-response and multiple-choice items", 28 *J. Educational Measurement*, 77-92 (1991). See for a study of the format problem with tests on the ability level of analytic thinking: B. Bridgeman and D. A. Rock, "Relationships among multiple-choice and open-ended analytical questions", 40 *J. Educational Measurement*, 313-329 (1993).

7 For a more extensive description of the curriculum see: J. H. C. Moust, "The Problem-based Education Approach at the Maastricht Law School", 32 *The Law Teacher*, 5-37 (1998).

ass
cri

V
ope
by
que
of t
cou
effi

In 1
star
of th
kno
kno

8 I
C
C
N
a
S
9 S
10 I
r
h
w
11 I
A

Both within and without the faculty, criticism grew against an assessment system only consisting of all-essay tests.⁸ The main points of criticism were:

- a. An all-essay test allows a limited number of questions only. Therefore, the content coverage is limited. This stimulates an undesirable study strategy. Students were able to pass their test, after studying only part of the subject matter.⁹ Potential gaps in their knowledge base failed to become manifest. The limited content coverage of the test also negatively affected the discussions in the tutorial groups. Students tended to ignore topics, which they expected not to be tested.¹⁰
- b. Due to growing student numbers, the marking of the tests placed a heavy burden on faculty members. Currently, the annual intake amounts to 550 students.
- c. The large number of teachers involved in marking, often more than 15 teachers for one test, and their very different backgrounds resulted in substantial marking variations.

Van Berkel and Crombag¹¹ claimed that an achievement test with both open and closed questions could offer a solution to the problems caused by the essay test. They argued that a test of both open and closed question formats would contain more questions, allowing better coverage of the subject matter. It would no longer pay to study only part of the course content. It was to be expected that such a test would be more efficient and less unreliable.

3. The Combination Test Experiment

In 1995, an experiment was conducted with the "combination test". It started as an experiment, because faculty wanted to ensure the quality of the tests. The scores on the multiple-choice questions had to measure knowledge of the law, skills and insight, not only factual legal knowledge.

8 Both students and faculty criticised aspects of our assessment system that was mainly based on tests consisting of four to six essay questions. This criticism was supported by the findings of external quality assessments in 1991 and 1997 by the Association of Universities in the Netherlands (VSNU). Since 1988, the Netherlands has a national system for external quality assurance in education. An external committee assesses the state of affairs in the discipline as a whole in all faculties and reports its findings. The report is made public and has a substantial impact on public opinion as to the quality of education at the different faculties.

9 Students avoided in particular such technical topics as alimony calculation.

10 In 1989, a student described how this worked. The student who starts a discussion on a theme not directly related to the expected test content is corrected by the other students in the tutorial group. Such a student receives comments such as: "What is this? They won't ask this in the test." R. Wolleswinkel, "Studying in a problem-based learning program", *Experiences with problem-based learning*, 30-31, Deventer (1989).

11 H. J. M. van Berkel and H. F. M. Crombag, *Essay questions versus multiple-choice questions: A research proposal*, (1992).

Many of the lecturers were suspicious of the use of a combination of multiple-choice and open-ended questions for testing. Their main questions were: "What is the validity of a combination test? How does a combination test affect the study behaviour of the students? Is a combination test more efficient than an essay test? How do students assess different test forms?"

To answer these questions, information was gathered from the results of a number of pilot combination tests. In 1995, an achievement test consisting of 40 four-option multiple-choice questions and one extensive essay question was used to evaluate a first-year course. In 1996, two first-year courses were evaluated by similar tests.

The test questions were drafted by 12 faculty members. These followed a course in question drafting. Before a question was included in the test, it was screened by a test review committee, made up from experienced teachers from the different legal domains, and an educationalist.

The most important screening criteria were:

- a. Relevance. This is especially important in problem-based learning, since students enjoy a relative freedom in their choice of subject matter. In consequence, the questions need to deal with the essential elements of the subject matter which all students need to study.
- b. Specificity. Can the question be answered on the basis of specific knowledge content rather than on the basis of general knowledge?
- c. Objectivity, degree of difficulty and discriminating power. Is the question unequivocal and unambiguous? Is a first year student able to answer the question? Is the question neither too easy nor too difficult for first-year students?¹²

The three tests were given to all first-year students. They were mandatory and the students had to pass all of them in order to go on to the second year. Afterwards, the question scores were analysed as to the degree of difficulty and discriminating power. The test review committee discussed the question statistics. After taking the test, the students were allowed to take the test questions home and critique them within 10 days. Where the comment of students or the statistical results of the question analyses gave cause for doubting the content of the question, it was deleted and the scores were corrected.

12 Questions which all students answer correctly, or all miss, do not contribute anything to the differentiation between high and low achieving students. Other factors influencing the ability of a question to discriminate, in addition to the degree of difficulty, are the extent to which there is only one correct answer and the attractiveness of the wrong answers to students of less ability, the soundness of the idea underlying the questions and the clarity of its expression. R. L. Ebel and D. A. Frisbie, *Essentials of Educational Measurement*, 5d ed., 223-224, New Jersey (1991).

dr
an
to
tes

4.1
"V
ou
cor
lev
cho
at
me
cor.
on
essa
knc
sco
A l
If t
Stu
the
for
corr

the
perf
Both
perf
perf

13 T
v
n
c
a
o
s.
a
e
A
14 C
v.
+
oi
(I

To estimate the drafting and marking time involved, the question drafters and markers were asked to record the time spent in drafting and marking. Immediately after taking the test, the students were asked to complete a questionnaire on the test itself, their preparation for the test and the influence of the test form on their study behaviour.

4. The Validity of a Combination Test

4.1 Method

"Validity" relates to the appropriateness of the interpretation of the outcome of the test in connection with its purpose.¹³ In the case of the combination test, we intend to measure legal knowledge mainly at the level of insight, application and problem-solving skills. The multiple-choice questions and the essay questions should measure legal knowledge at this (higher) cognitive level. Therefore both question forms must measure the same legal knowledge. Information on the validity of a combination test is found by comparing first-year student performance on a combination test with the performance of the same students on all-essay tests. If both all-essay and combination tests measure the same knowledge, student performance on both types of test is similar; students scoring high on combination tests will also score high on all-essay tests. A low-scoring student will score low, it is assumed, in both test formats. If this is the case, both test formats measure the same knowledge. Student performance is compared by calculating the correlation between the scores on the combination test(s) and the all-essay tests. If both test formats measure the same knowledge, the scores have a strong correlation.¹⁴

The validity of a combination test can also be examined by comparing the performance of a student on the multiple-choice questions with his performance on the essay question(s) within a single combination test. Both forms intend to measure the same knowledge. A student who performs well in the case of the multiple-choice questions is expected to perform well on the essay question(s) also. The correlation between the

13 The concept of "test validity" has changed in the course of the twentieth century. Traditionally, validity was regarded as a test characteristic and defined as the extent to which a test actually measures what the test makers intended to measure. Therefore, validity was seen as a constant. More recently, the ideas on test validity have changed and nowadays validity is associated with the use of test scores rather than with the test itself. For example, the scores of a test can be valid for assessing freshmen's knowledge of constitutional law. But when the same test is used to assess final-year students' knowledge of constitutional law, the test scores are less valid and should therefore not be used for selecting purposes. For a more extensive explanation of test validity, we refer to R. L. Linn and N. E. Gronlund, *Measurement and Assessment in Teaching*, 7d ed., 47-80, New Jersey (1995).

14 Coherence is expressed by a correlation coefficient: value +1 maximum positive coherence, value zero no coherence and value -1 maximum negative coherence. Perfect correlations of +1 or -1 are rarely found. Most coefficients of correlation in social research are around +0.50 or less. L. Cohen and L. Manion, *Research Methods in Education*, 4d ed., 130-132, London (1994).

multiple-choice scores and the essay question scores within a single combination test was also calculated. If a high correlation was found, it may be assumed that both question forms of the combination test measure the same knowledge.

4.2 Results

Table 1 shows the correlation between the scores on the multiple-choice questions and the scores on the essay question(s) in the combination tests.

Table 1. The correlation between the scores on the multiple-choice questions and the scores on the essay question(s) in the combination tests

Combination Tests	Students (N)	Pearson Correlation
April 1996	493	.51*
January 1997	558	.58*
April 1997	496	.60*

* $p < 0.01$.

The scores on the multiple-choice questions and on the essay questions within a combination test show quite a strong and statistically significant coherence, especially since the combination tests of April 96 and April 97 contained one essay question only and the January 97 test had two essay questions.¹⁵ The small number of essay questions per combination test has a downward effect on the correlations. The subject of the one or two essay questions has a strong influence on student performance. Therefore, in this case, the correlations can never reach the maximum value of +1. In this context, we have to establish that the correlations found are high. We can conclude that students who scored high on the multiple-choice questions also achieved a high score on the essay question(s) of the same test.

The multiple-choice questions were also correlated with the all-essay tests of the first year. The all-essay tests contain a larger sample of essay questions than the one or two essay questions of the combination tests. Therefore, the question subjects of the all-essay test have smaller

¹⁵ The correlation coefficients are statistically significant with a probability value of 0.01. This means that we are 99% certain that there is a statistical relation between the scores on both question forms. In social sciences, we must accept the small probability that there is no such relation.

inf
cor

of

Tal

—

Mu
Qu

—

Apr
Jan
Apr

* $p <$

que:
All
both

5

5.1 A

Imm
ques
their
stud
know
form
coml
studi
comb
ques
a bro
could
agree
some
Th
they

influence on student performance. This is represented by the higher correlations in Table 2.

Table 2 shows the correlation between the multiple-choice questions of the combination test and the all-essay tests of the first year.

Table 2. The correlation between multiple-choice questions of the combination tests and the first-year all-essay tests

Multiple-choice Questions	All-essay tests				
	Oct 95	Jan 96	July 96	Oct 96	July 97
April 1996	.78*	.80*	.85*		
January 1997				.70*	.73*
April 1997				.71*	.80*

* $p < 0.01$

We find high correlation coefficients between the multiple-choice questions of the combination test and the all-essay tests of the first year. All the correlations are statistically significant. This is an indication that both question formats measure the same attributes.

5. The Influence of the Test Format on Student Study Behaviour

5.1 Method

Immediately after taking the test, the students were asked to fill in a questionnaire on, among other things, the influence of the test format on their study behaviour. For every variable (self-study time, breadth of study, deep versus surface learning, and focus on reproduction of knowledge) students were asked to rate the influence of the question format of the test on a five-point scale. The questionnaire for the combination test contained items such as: "During this course period, I studied the course content in a broader way, because the exam consists of a combination of multiple-choice and essay questions" and for the essay test, questions such as "During this course period, I studied the course content in a broader way, because the exam consists of essay test questions." Students could give a score on a scale from 1 (full disagreement) to 5 (full agreement). Values 2, 3 and 4, stand for some disagreement, neutral and some agreement, respectively.

The students were also asked to estimate the weekly amount of time they spent studying for the course.

5.2 Results

The data on the question of format influence is reported in Table 3.

Table 3. Mean scores on the variables of the influence of question format on study behaviour (combi-test = combination test)

	Combi-test 1		Combi-test 2		All-essay test	
	Mean	N	Mean	N	Mean	N
Self-study time	2.9	318	2.9	297	3.0	324
Broad learning	3.0	320	3.0	302	2.9	330
Deep learning	3.0	321	3.0	300	3.0	325
Surface learning	2.3	317	2.4	301	2.3	326
Reproduction learning	2.8	320	2.7	302	2.8	322

In the students' perception the question format had no influence on their study behaviour and on the hours of self-study. The mean scores for all three tests are nearly similar. For the study behaviours which are considered positive in higher education—more self-study time, broad and deep learning—the mean scores are neutral. Study behaviour which is considered less desirable—surface learning—scored below neutral. Reproduction learning is rated a little below neutral.

All the students were asked to estimate how much time per week they studied. The mean hours per week are shown in Table 4.

Table 4. Amount of student-estimated self-study time in mean hours per week (combi-test = combination test): Mean and Standard Deviation (SD)

Tests	<i>All-essay</i>			<i>Combi-test</i>			<i>t</i> values
	Mean	SD	(Year)	Mean	SD	(Year)	
Test 2	14.1	6.4	(94/95)	15.8	9.2	(96/97)	2.79***
	13.9	7.1	(95/96)	15.8	9.2	(96/97)	3.13***
Test 3	14.1	7.1	(94/95)	16.2	9.2	(95/96)	2.47**
	14.1	7.1	(94/95)	15.9	8.1	(96/97)	2.22*

* $p < .025$. ** $p < .01$. *** $p < .005$.

It is remarkable that the proportion of self-study hours increases significantly in those years in which a combination test is used to examine students. In the academic year 95/96, the combination test was used for the first time and self-study time increased by 1.6 hours per

we
con
bo
of
stu

6.1
Fac
two
eac
for
on

6.2
In 1
mar
in t

Tab

—
Yea

—
Dra
Mar

—
Tota

* $p <$

N
choi
for c
yielc
T
essay

week. In 96/97, a second course was evaluated with the aid of a combination test for the first time; the study-time rose by 1.9 hour. In both cases, the content and format of the courses were similar to those of the preceding years. The statistical significance of the increase of self-study hours is a result of the change in test format and not of chance.

6. Are Combination Tests More Efficient Than Essay Tests?

6.1 Method

Faculty members were asked to indicate the hours spent in drafting the two types of questions and marking the answers on essay questions. In each year, this was done for one or two combination test(s) as well as for one all-essay test. Thus, it was possible to compare both test formats on efficiency.

6.2 Results

In Tables 5 and 6 the hours are shown which were spent on drafting and marking the tests. Table 5 relates to an all-essay and a combination test in the year 95/96.

Table 5. Drafting and marking time in hours for both essay and combination tests in the academic year 95/96

Year 95/96	Test 2 (all-essay)	Test 3 (combi)	<i>t</i> values
Drafting	10 hours	71 hours	1.67*
Marking	321 hours	80 hours	2.82**
Total	331 hours	151 hours	1.62*

* $p < .10$. ** $p < .005$.

Notwithstanding the additional effort required in drafting 40 multiple-choice questions and one essay-question (seven times the time needed for drafting the four questions of the all-essay test) the combination test yielded a significant saving of 54% (180 hours).

The time used to draft and mark two combination tests and one all-essay test in the year 96/97 is given in Tables 6a and 6b.

Table 6a. Drafting and marking time in hours for both test 4, an all-essay test, and test 2, a combination test, in the year 96/97

Year 96/97	Test 2 (<i>Combi</i>)	Test 4 (<i>All-essay</i>)	<i>t</i> values
Drafting	89 hours	18 hours	2.00*
Marking	161 hours	366 hours	2.69**
Total	250 hours	384 hours	0.22

p* < .05. *p* < .01.

Table 6b. Drafting and marking time in hours for both test 4, an all-essay test, and test 3, a combination test in the year 96/97

Year 96/97	Test 3 (<i>Combi</i>)	Test 4 (<i>All-essay</i>)	<i>t</i> values
Drafting	53 hours	18 hours	3.00**
Marking	71 hours	366 hours	2.26*
Total	124 hours	384 hours	1.04

p* < .025. *p* < .01.

In 96/97, the combination test also saved a substantial amount of time for faculty members. Test 2 yielded a reduction of 35% (134 hours) and for test 3 the reduction was 68% (260 hours).¹⁶ The relatively lower time-saving of the second test, compared to the third test, can be explained by the fact that test 2 contained two essay questions, whereas test 3 contained a single essay question. Furthermore, in the year 95/96, the question drafters of test 3 gained experience in drafting multiple-choice questions. The subsequent year, they became more efficient: 53 hours for question drafting instead of 71 hours. The question drafters of test 3 were more efficient than their colleagues of test 2. Most question drafters of test 2 had no experience in drafting multiple-choice questions. There is a clear learning effect.

¹⁶ Because of the small number of teachers involved in drafting and marking the combination tests the mean differences are not statistically significant. To obtain statistically significant results, larger samples are required. In the experiment all the teachers involved in the test were questioned. The "sample" is not really a sample, because it is equivalent to the whole population. It is the maximum sample size and cannot be increased. However, it is evident that the combination tests yielded considerable time savings.

7.1

In
dif
the
(b)
40

7.3

In

Tab

—
Tes

—
Apr
Apr
June
Oct

**p* <

Tab

—
Test

—
Apr
Apr
June
Oct

**p* <

T
prefe
test.
ques

7. Students' Preferences for Test Formats

7.1 Method

In the questionnaire, the students were also asked to assess three different test formats on a five-point scale. The three test formats were the following: (a) a test consisting of multiple-choice questions only, (b) a test consisting of essay questions only and (c) a test consisting of 40 multiple-choice questions and one or two essay questions.

7.3 Results

In Tables 7a and 7b the students' preferences for test formats are shown.

Table 7a. Students' preferences for the all-essay test format and the combination test format

Tests	<i>All-essay</i>		<i>Combi-test</i>		<i>t</i> values
	Mean	SD	Mean	SD	
April 96 (Combi)	3.0	1.3	3.5	1.2	5.24*
April 97 (Combi)	2.7	1.4	3.5	1.2	7.46*
June 97 (All-essay)	3.0	1.3	3.5	1.1	5.27*
October 97 (Combi)	2.5	1.3	3.8	1.1	14.40*

* $p < .0005$.

Table 7b. Students' preferences for the multiple-choice test format and the combination test format

Tests	<i>Multiple-choice</i>		<i>Combi-test</i>		<i>t</i> values
	Mean	SD	Mean	SD	
April 96 (Combi)	2.8	1.3	3.5	1.2	7.34*
April 97 (Combi)	2.7	1.3	3.5	1.2	7.78*
June 97 (All-essay)	2.5	1.3	3.5	1.1	11.02*
October 97 (Combi)	2.7	1.3	3.8	1.1	12.19*

* $p < .0005$.

The majority of students had a clear and statistically significant preference for a combination test over an all-essay or multiple choice test. Tests consisting of multiple-choice questions only or of essay questions only were less popular.

8. Conclusions

A test combining multiple-choice questions and essay questions should offer a solution to the problems that arise when a large student population is to be tested.

The results of this experiment show that there is a strong relation between the scores on essay questions and the scores on the multiple-choice questions. This implies that it is possible to use multiple-choice questions to assess knowledge of the law, insight and skills at a higher cognitive level. This fully bears out the conclusion of an earlier study in legal education that both open and closed questions measure the same construct.¹⁷

Students do not list differences in study behaviour as a result of different test formats. At the same time, students indicate that they study more hours when the course is evaluated by a combination test. There are two possible explanations for the increase in study hours:

- (a) The test expectation effect. When preparing for an achievement test, students have certain ideas on how they will have to reproduce their knowledge during the test. As regards multiple-choice questions, they expect that a more detailed knowledge is required, whereas for essay questions they expect that a more global knowledge of the correlations within the subject matter may be essential. For a combination test, students expect that an appeal will be made to both kinds of knowledge. The probable result is that students spend more hours on test preparation.
- (b) The acquaintance effect. If a certain course has a fixed content and comes with the same assessment procedure over the years, old test questions are likely to circulate widely among students. The better acquainted with the content and format of the questions, the more efficient the student's preparation for the achievement test. This will translate into less self-study time. The combination of different test formats is a novelty for the students, not experienced earlier in their law studies. This will heighten their feeling of insecurity. The students will try to reduce these feelings by studying longer for the innovative test.

Compared to all-essay tests, combination tests yield considerable savings in faculty input. The time needed in drafting the questions for a combination test and marking the answers amounted to two-thirds or less the time required for drafting and marking an all-essay test. In

¹⁷ The researcher worded his conclusion as follows:

No matter how we look at it, there is no single indication that a test with open questions measures something else than an objective test.

H. F. M., Crombag, "Comparison between an open-question examination and a multiple-choice test", 25 *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, at 358-359 (1970).

com
mu
me
in
Fu
eff

po:
inc
inil
cho
acc
has
con
hav
que
The
yea
firs
test
enti

usir
bro.
cho
eva
forn
stuc

])
to n
leve
essa
coac
adr

contradistinction to what law teachers tend to think, the drafting of multiple-choice questions proved not to be a problem for faculty members provided they received advance instruction. The vast savings in time for marking greatly outweighed the additional drafting required. Furthermore, drafting experience contributed considerably to drafting efficiency.

The time required for drafting multiple-choice questions proved a point of concern to some faculty members. Actual time savings would increase the acceptance level of the combination test. In spite of the initial distrust by many faculty members of a test containing multiple-choice questions the combination test has now gained faculty-wide acceptance. The faculty members involved feel that the combination test has improved the quality of our student assessment system and has contributed considerably to the lessening of faculty workload. Students have also expressed a distinct preference for a test combining different question formats over a test in which just one question format is used. The positive results achieved with the combination test in the academic year 97/98 has prompted the faculty board to implement this test for all first-year courses. From the academic year 98/99 onwards, combination tests are used to assess the students in the compulsory courses of the entire law curriculum.

The conclusion of this experiment is unambiguously in favour of using combination tests, and combining the best of both worlds: meaning broad content coverage, efficiency and fairness in marking the multiple-choice questions, combined with measuring creativity, writing skills or evaluative skills through essay questions. Combining the two question formats in a single test results in higher study input on the part of the students and a high acceptance by both students and faculty.

If the question drafters take their task seriously, it makes it possible to measure knowledge of the law, insight and skills on a higher cognitive level through a test that consists of both multiple-choice questions and essay questions. In consequence, all novice question drafters need to be coached and the questions must be critically reviewed before the test is administered.

ould
lenttion
iple-
oice
gher
ly in
ameIt of
tudy
Therement
ve to
tiple-
lge is
more
natter
at an
able
n.ontent
years,
dents.
of the
or the
time.or the
is will
try to
e test.
derable
is for a
irds or
test. In

questions

multiple-
359 (1970).

The **LAW TEACHER**



Association of Law Teachers

THE
INTERNATIONAL JOURNAL
OF LEGAL EDUCATION



1999
Volume 33
Number 2