Advances in Health Sciences Education (2006) 11:155–171
DOI 10.1007/s10459-005-5290-9


# Differential Effects of Two Types of Formative Assessment in Predicting Performance of First-year Medical Students★

SALLY KRASNE[1,*], PAUL F. WIMMERS[2], ANJU RELAN[3] and THOMAS A. DRAKE[4]

[1]*Department of Physiology, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095-1751, USA;* [2]*University of Pittsburgh School of Medicine, Pittsburgh;* [3]*IDTU, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA;* [4]*Department of Pathology, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA (*author for correspondence, Phone: +310-825-8884; Fax: +310-206-3788; E-mail: skrasne@mednet.ucla.edu)*

(Received 7 December 2004; accepted 18 November 2005)

**Abstract.** Formative assessments are systematically designed instructional interventions to assess and provide feedback on students' strengths and weaknesses in the course of teaching and learning. Despite their known benefits to student attitudes and learning, medical school curricula have been slow to integrate such assessments into the curriculum. This study investigates how performance on two different modes of formative assessment relate to each other and to performance on summative assessments in an integrated, medical-school environment. Two types of formative assessment were administered to 146 first-year medical students each week over 8 weeks: a timed, closed-book component to assess factual recall and image recognition, and an un-timed, open-book component to assess higher order reasoning including the ability to identify and access appropriate resources and to integrate and apply knowledge. Analogous summative assessments were administered in the ninth week. Models relating formative and summative assessment performance were tested using Structural Equation Modeling. Two latent variables underlying achievement on formative and summative assessments could be identified; a "formative-assessment factor" and a "summative-assessment factor," with the former predicting the latter. A latent variable underlying achievement on open-book formative assessments was highly predictive of achievement on both open- and closed-book summative assessments, whereas a latent variable underlying closed-book assessments only predicted performance on the closed-book summative assessment. Formative assessments can be used as effective predictive tools of summative performance in medical school. Open-book, un-timed assessments of higher order processes appeared to be better predictors of overall summative performance than closed-book, timed assessments of factual recall and image recognition.

**Key words:** assessment/testing, curriculum development/evaluation, educational measurement, on-line testing, student evaluation, teaching methods, undergraduate medical education

★This research was presented at the 86th Annual Meeting of the American Educational Research Association (AERA) in Montreal, Canada, April 11–15, 2005.

Improving the quality of high stakes assessment for undergraduate medical student training has invariably evoked the interest of medical educators (Rolfe and McPherson, 1995). In contrast to "summative" assessments, implemented to make decisions about student progress or certification, little attention has been devoted to "formative" assessments, which are undertaken to help develop students' intellectual capabilities for improved achievement, to identify and act upon students' strengths and weaknesses, and to modify teaching practices if needed (Mennin and Kalishman, 1998). Formative assessments are systematically designed interventions administered during the assigned period of instruction to enhance cognitive and motivational support for learning and achievement in multiple ways: by informing students of the gap between their existing and expected knowledge states, by familiarizing students with what will be expected on summative assessments, by providing feedback on students' thinking, and by guiding the direction of their learning. Because they are not part of summative assessment, these instructional interventions occur within a non-threatening environment (Rolfe and McPherson, 1995; Sadler, 1989; Wass et al., 2001), which may account for the positive response such assessments elicit among students across all types of formats and domains (Henly, 2003; Hill et al., 1994; Houghton and Wall, 2000; Paschal, 2002; Peat and Franklin, 2002; Ritter, 2000; Vaz et al., 1996; Velan et al., 2003). A recent report published by the Association of American Medical Colleges (2004) proposing curricular reform guidelines towards an ideal medical education system states that such a system will, "conduct a rigorous assessment of learners' abilities throughout the course of their careers, to assist them in improving their performance (formative assessment) and to ensure that they have achieved the level of performance required to advance professionally (summative assessment)."

An overriding purpose of developing and administering formative assessments is to make a measurable difference in student achievement on summative assessment. As pointed out by Ramsden (1992), the assessment *is* the curriculum, and by-and-large, students adjust their learning to what they expect to be tested on. Thus, formative assessments can serve as excellent tools for shaping the way students approach the material in a course. In addition, it has long been known (Ebbinghaus, 1885/1913) that greater retention of knowledge occurs when learning trials are spaced rather than massed, and weekly formative assessments are likely to promote distributed studying throughout the period of the course rather than just before the final exam. Given the large amount of faculty effort and time involved in planning, developing and administering formative assessments, integrating these into the curriculum must be evaluated against competing instructional strategies to achieve the same goals.

Several studies have suggested that formative assessments can improve performance on subsequent summative assessment (Bondemark et al., 2004; Greer, 2001; Thiesen-Roe et al., 2004). These studies have examined the effects of formative assessments on achievement via inferential, correlational analyses based on a linear causal framework. In this paper, we hypothesize that complex, interactive relationships may exist among multiple formative assessments and performance which warrant an investigation. The purpose of the present study is to examine, for beginning medical-students, the relationships between performance on required, weekly, formative assessments and on summative assessments. These assessments were given within the context of an integrated curriculum in which each week's material applied multiple disciplines to understanding the topic being presented. For the curricular block studied in this report, basic science concepts, application of basic science to clinical problems, histopathology and superficial anatomy, and concepts underlying interviewing and physical examination were all applied to understanding basic processes underlying disease. Within this curricular context, we investigated how achievement on formative assessments predicts performance on a summative assessment as a whole, as well as by type of assessment. Assessment formats included a timed, closed-book assessment aimed at evaluating factual recall and image recognition (Peitzman et al., 1990) and referred to here simply as Closed-Book, and an untimed, open-book/resource assessment designed to evaluate higher order skills (Peitzman et al., 1990) including the ability to synthesize and apply factual knowledge to complex questions as well as to effectively utilize resources, referred to here simply as Open-Book. This study further seeks to determine the interrelationships between performance on the two types of assessments themselves.

The following specific questions are addressed by the present research: (1) How does performance on formative assessments relate to performance on summative assessments? (2) What is the relationship between the Open-Book and Closed-Book components of the assessments? (3) What is the relationship between performance on the Open-Book vs. Closed-Book components of the formative assessments and performance on the summative assessment?

## Methods

Seven required and one optional formative assessment, corresponding to weekly curricular ''themes'' and consisting of both Open-Book and Closed-Book components, were delivered on-line over the 8-week curricular block. Topics covered from the first through eighth weeks are presented in Table I. All students were required to take the formative assessments for weeks 1–7

*Table I.* Block 1 weekly structure and associated assessments

| Wk | Topic | Closed-Book assessment Mean ± SD % | Open-Book assessment Mean ± SD % | CB&OB Mean(±SD) % |
|---|---|---|---|---|
| 1 | Genetics and molecular biology | 67.2 ± 11.5 | 89.7 ± 11.9 | 78.4 |
| 2 | Embryology and signal transduction | 78.5 ± 10.3 | 77.8 ± 16 | 78.1 |
| 3 | Surface anatomy and cell biology | 81.7 ± 9.1 | 74.6 ± 15.3 | 78.1 |
| 4 | Neoplasia | 77.2 ± 10.5 | 81.5 ± 14.6 | 79.3 |
| 5 | Electrophysiology, cell homeostasis and cell injury | 80.0 ± 12.6 | 74.5 ± 15.5 | 77.2 |
| 6 | Inflammation | 83.0 ± 8.1 | 83.0 ± 10.1 | 83.0 |
| 7 | Immunology | 85.2 ± 9.3 | 86.4 ± 12.0 | 86.8 |
| 8 | Pharmacodynamics and therapeutics (optional assessment) | Not scored | Not scored | Not scored |
| | *Formative assessments 1–7* | *79.0 ± 5.8* | *81.1 ± 7.6* | *80.1 ± 5.7* |
| 9 | *Final exam (summative)* | *84.3 ± 7.4* | *84.0 ± 7.04* | *84.1 ± 6.1* |

whereas that for week 8 was optional to prevent overloading students immediately before the summative assessment. The assessment for week 8 was excluded from the analysis as it represented incomplete data. For weeks 1–7, the Closed-Book assessments ranged from 16 to 30 questions (16–50 points), and the Open-Book assessments ranged from 8 to 14 questions (20–33 points). Scores were calculated as percentages of total points on each part of the assessment. A summative assessment containing both Closed-Book and Open-Book components, of 84 and 48 questions, respectively, was delivered in the ninth week.

Both Closed-Book and Open-Book assessments consisted of multiple-choice and fill-in-the-blank questions whereas Open-Book assessments additionally included short-answer questions. Examples of questions asked on Closed-Book and Open-Book formative and summative assessments are given in Appendix 1. Assessments were delivered online via the "ANGEL" course management system (http://www.angellearning.com/). Formative assessments could be taken anytime and place between Friday evening and Monday morning following the particular week. All multiple-choice and fill-in-the-blank questions were scored electronically, whereas short-answer

questions were scored online by Problem-based Learning session tutors for the typically eight students in their groups based upon a scoring template suited to each question. Timing of the Closed-Book component was done within ANGEL's quiz-function, and Closed-Book assessments were automatically submitted at the end of the allotted time if they had not been submitted previously. ANGEL also tracked all online activity within the course for each student. Feedback on electronically scored questions was delivered upon completion of the assessment and consisted of correctness of responses and sometimes a brief explanation for incorrect responses. Feedback on the short-answer questions always occurred upon completion of the assessment, whereas the scores for these types of questions were delivered by Tuesday noon following the weekend. The summative Open-Book component was taken over a 3-day period at any location, while the proctored, Closed-Book component was taken over a 3-hour period at the medical school. All 146 students enrolled in the course took all seven required formative assessments and the summative assessment.

## Statistical analysis

Analysis was implemented in two successive steps. The first, exploratory step of data analysis was carried out using SPSS version 12.01 for Windows. The statistical procedures included the student *t*-test, to calculate the differences between the means of the different assessments, and correlation coefficients to determine the relationships between all observed variables.

In the second step, several models were developed to examine the research questions posed in this study. The hypothesized models were analyzed and compared via Structural Equation Modeling (SEM) using the software package EQS 6.1 (Bentler, 2003). SEM models could reflect causal relationships between the assessment components and underlying latent constructs. These causal processes are represented by the regression equations. The relationships between the observed variables and their underlying latent constructs (factors) are defined in terms of weights, the path coefficients. The hypothesized models can be tested statistically in a simultaneous analysis of the entire system of variables to determine the extent to which it is consistent with the data (for a practical guide, see Byrne, 1994). The fit indices provided by the SEM package were used to determine whether the model adequately fit the data and are shown at the bottom of each figure. Bentler's Comparative Fit Index (CFI) compares the fit of the particular model under test with a model in which none of the variables are related; a CFI of 0.90 or higher has typically been taken as indicating good fit between the model tested and the data (a value of 0.95 has been proposed by Hu and Bentler, 1999, but has

been debated by Marsh et al., 2004). The Standardized Root Mean-Square Residual (SRMR) represents the average standardized discrepancy between observed and model-implied relations; a value below 0.08 indicates good fit. Finally, Steiger's Root Mean Square Error of Approximation (RMSEA) adjusts for a model's complexity. A value below 0.05 indicates proper fit.

## Results

### How does performance on the formative assessments relate to performance on the summative assessment?

In order to address the first research question, we initially compared mean scores of the seven required formative assessments with those of the summative assessments (see Table I) and found the scores on summative assessment to be significantly higher, $t(145) = 9.11$, $p < 0.001$. In addition, there was a statistically significant correlation between individuals' mean scores on formative assessments and summative assessments, $r = 0.58$, $p < 0.001$.

In exploring how performance on the formative assessments might be related to performance on the summative assessment, two alternative structural models were compared using SEM. The first, a serial model, assumes that successive contributions of formative assessments (from week 1 to 7) have led to the increase in summative assessment scores (Figure 1a). For example, it could be expected that incremental gains would arise from general, topic-independent learning such as familiarity with assessment format and improved use of learning strategies that would yield improved performance from one week to the next, despite the change in topics each week. The second hypothesized model assumes that there are latent variables that determine achievement on formative assessments (FA factor) and on summative assessments (SA factor) and that the FA factor positively influences the SA factor (Figure 1b). For example, one might hypothesize that some intrinsic characteristic of the students (e.g. motivation, initial knowledge level) might predict how they will perform on each assessment with very little change occurring from one week to the next. Clearly, the values of the fit indices of the first, serial, model indicate no fit whereas those of the second, latent variable, model indicate a proper fit.

These results suggest that performance on each formative assessment does not directly influence performance on the subsequent formative assessment but rather that a latent variable contributes to performance on formative assessments and summative assessments. This model shows a strong influence of the FA factor on the SA factor (normalized regression coefficient = 0.91)

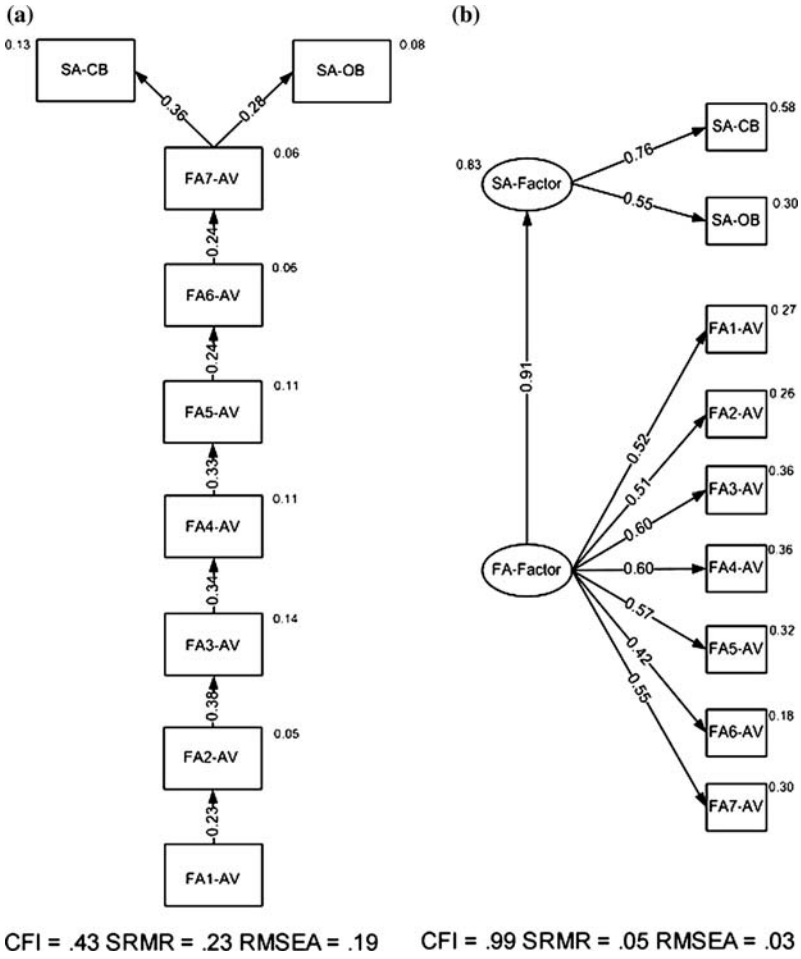



*Figure 1.* Two hypothesized models for the relationship between formative assessments and summative assessments. Figure 1a represents a model in which learning on each formative assessment (FA) contributes to performance on the subsequent assessment ultimately leading up to performance on the summative assessments (SA). Figure 1b represents a model in which performance on each formative assessment is independent of previous assessments but is, instead, determined by a latent factor; performance on the summative assessment is also determined by a latent factor, and the latent factor for formative assessment feeds into that for summative assessment. In all figures, normalized regression coefficients and correlation coefficients are shown as values within the arrows whereas the values adjacent to dependent variables and factors are normalized values of variance, $R^2$, and equal the proportion of the variable's variance that is explained by all of the variable's predictors. In this and subsequent figures, the number associated with the variable label indicates the week in which the formative assessment was given; AV indicates averages, CB refers to Closed-Book, and OB refers to Open-Book assessment scores in the indicated week. A correlation matrix with the means and standard deviations of all observed variables is presented in Appendix 2.

and suggests that improved performance on the summative assessment does not result simply because of improvement in some ability or knowledge with successive formative assessments. In such a model, the number of assessments is relatively unimportant, since linear, cumulative effects of formative assessments are not incorporated.

### What is the relationship between the open-book and closed-book components of the assessments?

There were no statistically significant differences between the mean scores of the seven Closed- and Open-Book formative assessments or between the means of the Closed- and Open-Book summative assessment (see Table 1). The correlations between the Closed-Book and Open-Book components of the formative assessments and between these components of the summative assessments were 0.48 ($p < 0.001$) and 0.41 ($p < 0.001$), respectively, indicating a moderate relationship between the average performance on Closed-Book and Open-Book components of each type of assessment. In order to explore whether individual latent factors exist that can predict performance on Closed-Book and Open-Book assessments and to what degree these factors, themselves, might be correlated, we applied the SEM model which was the best fit in Figure 1b to each component of the formative assessments. We then determined the correlation between the Closed-Book and Open-Book factors. This hypothesized model, as shown in Figure 2, fits the data well. The individual latent variables (FA-OB factor and FA-CB factor) identified for both Open-Book and Closed-Book components of formative assessments as represented in the model predict the scores on the Open-Book and Closed-Book assessments over weeks 1 through 7 to a moderately high degree, and the correlation between these two factors is higher than that derived from a simple correlation analysis which weights all assessments equally.

### What is the relationship between the latent factors deduced for closed-book and open-book formative assessments and performance on the summative assessment?

Last, we examined the degree to which the Open-Book and Closed-Book components of the formative assessments were predictive of the Open-Book and Closed-Book components of the summative assessments, as proposed in the third research question. Figure 3 shows the relationships between the FA-OB and FA-CB factors and their individual loadings onto the Open-Book and Closed-Book components of the summative assessment. This model, which demonstrates that FA-CB and FA-OB factors predict performance on

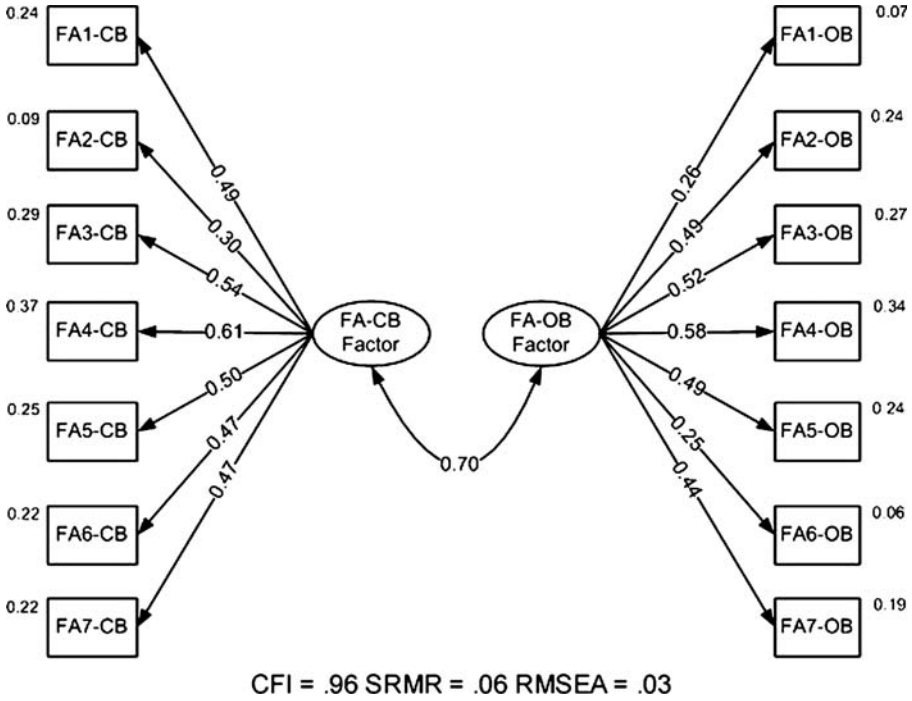



*Figure 2.* Correlation between latent variables "FA-CB Factor" and "FA-OB Factor" predicting Closed-Book and Open-Book formative assessments, respectively.

the summative assessment Closed-Book and Open-Book variables, also appears to fit the data well.

The model reveals that the Closed-Book and Open-Book summative assessment factors are well predicted by their formative assessment counterparts. In addition, based upon the relative loadings of FA-OB and FA-CB factors onto the two components of the summative assessment, the FA-OB factor has predictive value for both the Open-Book and Closed-Book components of the summative assessment, whereas the FA-CB factor appears to predict only the Closed-Book part of the summative assessment. These results indicate that the Open-Book formative assessments are much better predictors of overall performance on the summative assessment than are the Closed-Book formative assessments. Since the Closed-Book summative assessment latent variable has two moderately weighted inputs to it whereas the Open-Book has only one, a larger percent of the variance of performance on the Closed-Book summative assessment (50%) is predicted by the two formative assessment factors than is that of performance on the Open-Book summative assessment (30%).

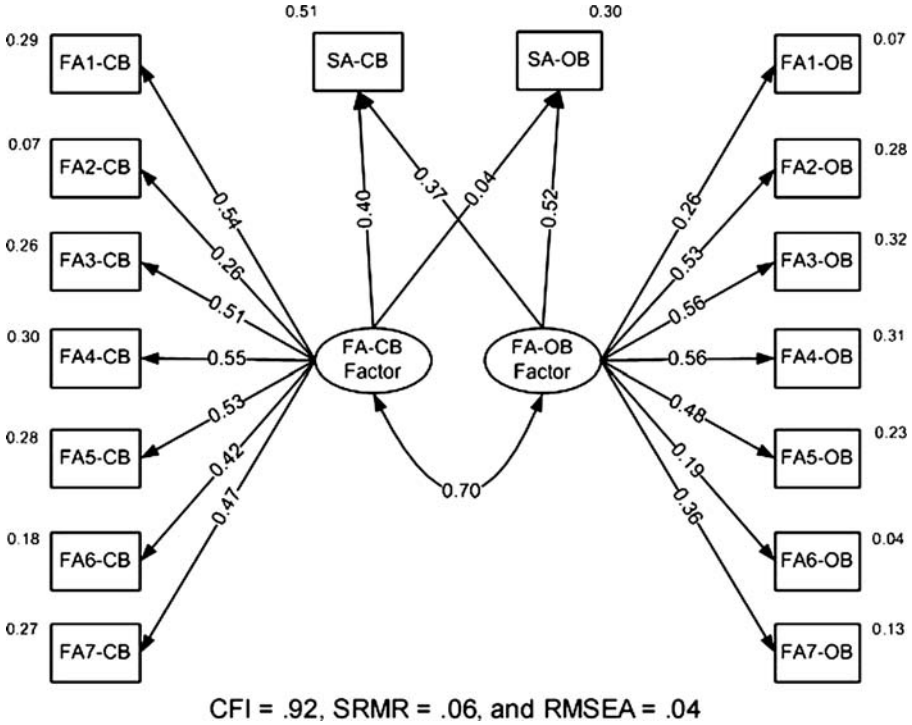



*Figure 3*. Relative contributions to Closed-Book and Open-Book summative assessments by the FA-CB and FA-OB factors for formative assessments.

## Discussion

The present study examined how performance on formative assessments relates to performance on a subsequent summative assessment and how performance on timed, Closed-Book assessments aimed at evaluating factual recall and image recognition (Closed-Book assessments) relates to performance on un-timed, Open-Book/resource assessments, designed to evaluate higher-order processes such as the ability to synthesize and apply factual knowledge to complex problems and effectively utilize resources in obtaining necessary information (Open-Book assessments). Structural Equation Modeling (SEM) allowed us to determine that a model in which latent factors predicted performance on formative and summative assessments fit the data well in contrast to an incremental model, in which learning from each formative assessment feeds into the subsequent one and ultimately into the summative assessment gave a poor fit to the data.

One of the main findings from this study is that different (but correlated) latent factors appear to underlie achievement on tests designed to assess factual recall and image recognition than on those designed to assess higher

order processes. The traits that these latent factors represent remain to be discovered; however, one might imagine several factors that could underlie different performance on these two forms of assessment (e.g. memorizing vs. ''deep learning,'' as designated by Marton and Saljo, 1976; differing degrees of motivation to access resources and solve the problems presented on the Open-Book assessment; different levels of anxiety evoked by the two formats). In trying to understand possible components of these latent factors, it is necessary to first consider the properties, other than question taxonomy, that differ between the two types of assessment. There are three such properties: the format of questions (multiple choice vs. a mixture of short-answer, multiple choice and fill-in-the-blank); access to materials (Closed-Book vs. Open-Book/open-web); and degree of time constraint (timed vs. un-timed).

The observed differences are unlikely to derive from the fact that some of the questions (approximately 25%) in the Open-Book assessments were in short-answer format. Studies have suggested that students use the same strategies in studying for and answering multiple-choice questions vs. short-answer or essay questions on timed, in-class assessments (Smith and Miller, 2005; Tang, 1992;). Other studies have shown that a short-answer vs. multiple-choice format, per se, does not appear to change the statistical or psychometric properties of exam performance once ''guessing'' on multiple-choice questions is accounted for (see Heck and Stout, 1998, and literature discussed therein). It is clear that different types of questions lend themselves to different formats; however, as discussed by Schurwith and Van der Vleuten (2003), the format of a question is not what determines what is tested by the question, and thus, it is unlikely that question format, *per se*, is as important a consideration as other possible differences between the two types of assessment.

The second property of the two test formats that differs is access to materials. The Open-Book assessments not only *allowed* access to materials but *required* such access in order to answer many of the questions. For example, the students might be asked to find a recently approved therapy for a particular autoimmune disease and to discuss its site(s) of action; so one set of abilities being examined in the Open-Book assessments was knowledge of which resources are most appropriate to locate different sorts of information and how to access and extract information from these resources. In our study, we did not find significant differences in the average performance on Closed-Book and Open-Book assessments so it is likely that the ''level of difficulty'' of the two types of assessment were comparable. However, studies have suggested that the Open-Book format reduces examination tension and stress and leads to lasting learning outcomes (Feldhusen, 1961; Jehu et al., 1970); Michaels and Kieran, 1973; Weber et al., 1983)

In terms of the third property, differing time constraints between the two forms of assessment, differences in strategy for multiple-choice tests vs. essays have been reported if the latter were done as ''take-home'' assignments, suggesting that the timed nature of an assessment may be an important variable (see Scouller, 1998). Because the Open-Book assessments in the present study were un-timed and taken at home, yet consisted primarily of multiple choice questions, it is not clear whether or not students utilized the same strategies as for in-class examinations. However, it is also worth considering that students might have used different strategies in approaching the two types of assessments because of the differences in types of information examined: factual recall and image recognition in Closed-Book assessments and abilities such as interrelating basic knowledge, applying basic information to more complex situations, accessing appropriate resources, and drawing conclusions from information, in the case of Open-Book assessments. Recently, Smith and Miller (2005) found that assessment ''type'' (multiple choice vs. essay questions) on timed, in class tests had no significant influence on how students approached their learning whereas the particular discipline did, for both assessment types. In the present study, it is possible that the types of tasks required by Open-Book, un-timed assessments require different strategies than Closed-Book, timed ones, especially since we used the two different formats to assess different types of knowledge and abilities.

A particularly important observation from the present study was that the latent variable underlying performance on Open-Book formative assessments appears to be more predictive of overall achievement on the summative assessments than does the latent variable underlying performance on Closed-Book formative assessments. It is interesting to consider why the Open-Book formative assessments would be much better at predicting outcome on the summative assessment than the Closed-Book formative assessments. One explanation might lie in the observations of Gay (1980) who compared the final exam scores on a selection of both short-answer and multiple-choice questions in an Introductory Educational Research course as a function of whether students took interim exams in short-answer or multiple-choice format. Although she observed no differences in performance on the interim exams as a function of format, she found that short-answer items appeared to lead to greater retention than multiple-choice items covering identical material. Thus something in the act of generating and writing the answer to a short-answer question compared to choosing among answers in multiple-choice questions may enhance retention of the material. The additional feature of the Open-Book format, allowing students to access knowledge not yet in their ready memory, may have provided additional reinforcement of their store of factual knowledge which could be carried forward to the

summative assessment as well. By contrast, no similar effect of the Closed-Book formative assessment on higher-order processes would seem likely.

The differential effects of the Closed-Book and Open-Book formative assessments in loading onto the summative assessments suggests that including a component of Open-Book/resource, un-timed assessment might provide added value over a strictly Closed-Book, timed format. Certainly, when doctors analyze cases, they need not restrict themselves solely to their ability to recall information, and as the knowledge base expands, doctors are becoming increasingly reliant on their abilities to efficiently and effectively utilize resources. Thus, a component of assessment aimed at accessing and applying knowledge would also promote an increased mastery of these skills.

As an initial attempt to unravel the interrelationships between assessments and performance, this study generates several questions for future research. To the extent that the assessments, themselves, test a range of intellectual modalities from visual and verbal memory to problem-solving to accessing literature, it seems unlikely that the latent variables represent single, identifiable constructs. Nothing in the present research can elucidate the degree to which aptitude, motivation, anxiety, or other qualities are embodied in the latent variables, FA factor and SA factor. Of particular interest, therefore, would be to identify cognitive and non-cognitive factors that influence performance on formative-assessments and the processes that contribute to improved performance on summative assessment.

From a long-term standpoint, understanding how formative assessments affect learning strategies would be of interest, as well as their predictive validity for performance on medical licensing exams or other high stakes summative assessments. Long-term retention resulting from Open-Book vs. Closed-Book assessments would shed light on the stability of positive effects of Open-Book assessments revealed in this study. Finally, the present study was conducted in the context of an integrated curriculum with beginning medical students, using unique content for each week. More varied instructional contexts and content in the study of formative assessments would help establish generalizable findings to guide practice.

**Acknowledgements**

The authors would like to thank Drs LuAnn Wilkerson and Gretchen Guiton for reading and providing helpful feedback on this paper.

## Appendix 1

Examples of questions on Closed-Book and Open-Book formative assessments

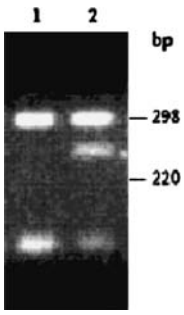
*Closed-Book (Week 2): Embryology and Signal Transduction*

While doing an ob/gyn residency, you deliver a pair of twins. You notice that they share a common placenta and chorionic sac but have separate amnionic cavities. What is the most likely origin of their twinning?

A. These are dizygotic twins
B. These are monozygotic twins that split at the two-cell stage.
C. These are monozygotic twins formed from splitting of the inner cell mass early in development.
D. These are monozygotic twins formed from splitting of the inner cell mass at a late stage of development

*Open-Book (Week 2)*

Another group studying this syndrome [spontaneous ovarian hyperstimulation syndrome] identified a mutation in the FSHr designated D567N. RFLP analysis was performed on a PCR amplified portion of exon 10, comparing an affected individual with an unaffected person. Details of this assay are as follows, and the results are shown in the figure below. PCR was used for specific amplification of a 530-bp segment of exon 10 of the follicle-stimulating hormone receptor centered on the mutation. Tsp45I cleaves twice the 530-bp PCR product obtained from the DNA of control subjects, generating one fragment of 290 bp and two fragments of 120 bp each (lane 1). The D567N mutation destroys the second Tsp45I restriction site, thus generating a mutation-specific band at 240 bp.




Based upon the RFLP pattern in lane 2 for the affected individual, is the mutation dominant or recessive? Justify your answer. (Not more than 2 sentences)

If the odds of carrying this mutation on both alleles are 1/1,000,000 in the population, what are the odds of being heterozygous for the mutation?

A. 1/50
B. 1/100
C. 1/250
D. 1/500
E. 1/1000

## Appendix 2

Means, Standard Deviations, and correlations of all observed variables

| *Observed variables* | Mean | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Students (N = 146)* | | | | | | | | | | | | | | | | | | |
| 1. FA OB 1 | 0.90 | 0.12 | – | 0.11 | 0.13 | 0.20* | 0.08 | −0.06 | 0.16 | 0.20* | −0.02 | 0.06 | 0.15 | 0.05 | 0.14 | 0.09 | 0.22** | 0.16 |
| 2. FA OB 2 | 0.78 | 0.16 | | – | 0.40** | 0.24** | 0.28** | 0.20* | 0.21* | 0.30** | 0.20* | 0.27** | 0.08 | 0.23** | 0.04 | 0.14 | 0.28** | 0.41** |
| 3. FA OB 3 | 0.75 | 0.15 | | | – | 0.34** | 0.18* | 0.08* | 0.23** | 0.24** | 0.08 | 0.24** | 0.05 | 0.15 | 0.10 | 0.16* | 0.32** | 0.38** |
| 4. FA OB 4 | 0.82 | 0.15 | | | | – | 0.30** | 0.11 | 0.27** | 0.18* | 0.11 | 0.11 | 0.28* | 0.09 | 0.33** | 0.23** | 0.35** | 0.31** |
| 5. FA OB 5 | 0.75 | 0.15 | | | | | – | 0.15 | 0.21* | 0.17* | 0.13 | 0.15 | 0.17* | 0.29** | 0.16 | 0.34** | 0.23** | 0.32* |
| 6. FA OB 6 | 0.83 | 0.10 | | | | | | – | 0.11 | 0.17* | 0.18* | 0.14 | 0.19* | 0.09 | −0.01 | 0.12 | 0.10 | 0.08 |
| 7. FA OB 7 | 0.86 | 0.12 | | | | | | | – | 0.17* | 0.01 | 0.25** | 0.17* | 0.14 | 0.14 | 0.17 | 0.10 | 0.19* |
| 8. FA CB 1 | 0.67 | 0.12 | | | | | | | | – | 0.08 | 0.32** | 0.28** | 0.31** | 0.22** | 0.16 | 0.21** | 0.43** |
| 9. FA CB 2 | 0.78 | 0.10 | | | | | | | | | – | 0.21** | 0.21* | 0.10 | 0.21* | 0.11 | 0.14 | 0.11 |
| 10. FA CB 3 | 0.82 | 0.09 | | | | | | | | | | – | 0.36** | 0.25** | 0.21* | 0.17* | 0.14 | 0.32** |
| 11. FA CB 4 | 0.77 | 0.10 | | | | | | | | | | | – | 0.31** | 0.40** | 0.29** | 0.16 | 0.31** |
| 12. FA CB 5 | 0.80 | 0.13 | | | | | | | | | | | | – | 0.19* | 0.25** | 0.07 | 0.44** |
| 13. FA CB 6 | 0.82 | 0.08 | | | | | | | | | | | | | – | 0.18* | 0.10* | 0.22** |
| 14. FA CB 7 | 0.86 | 0.09 | | | | | | | | | | | | | | – | 0.35** | 0.44** |
| 15. SA OB | 0.84 | 0.09 | | | | | | | | | | | | | | | – | 0.42** |
| 16. SA CB | 0.84 | 0.07 | | | | | | | | | | | | | | | | – |

*$p < 0.05$, **$p < 0.01$, two-tailed.

## References


AAMC (2004). *Educating Doctors to Provide High Quality Medical Care: A Vision for Medical Education in the United States* Washington, DC: AAMC Institute for Improving Medical Education.

Bentler, P. (2003). *EQS [Computer Software]. 6.1 Beta* Encino, CA: Multivariate Software, Inc.

Bondemark, L., Knutsson, K. & Brown, G. (2004). A self-directed summative examination in problem-based learning in dentistry: A new approach. *Medical Teacher* **26**(1): 46–51.

Byrne, B.M. (1994). *Structural Equation Modeling with EQS/Windows: Basic Concepts, Applications, and Programming* Thousand Oaks, Calif: Sage Publications.

Ebbinghaus, H. (1885/1913). *Memory: A Contribution to Experimental Psychology* New York: Teachers College, Columbia University.

Feldhusen, J.F. (1961). An evaluation of college students' reactions to open-book examinations. *Educational and Psychological Measurement* **21**: 637–645.

Gay, L.R. (1980). The comparative effects of multiple-choice versus short-answer tests on retention. *Journal of Educational Measurement* **17**(1): 45–50.

Greer, L. (2001). Does changing the method of assessment of a module improve the performance of a student? *Assessment & Evaluation in Higher Education* **26**(2): 127–138.

Heck, J.L. & Stout, D.E. (1998). Multiple-Choice vs. Open-Ended exam problems: Evidence of their impact on student performance in introductory finance. *Financial Practice and Education* **8**(1): 83–93.

Henly, D. (2003). Use of Web-based formative assessment to support student learning in a metabolism/nutrition unit. *European Journal of Dental Education* **7**: 116–122.

Hill, D.A., Guinea, A.I. & McCarthy, W.H. (1994). Formative assessment: a student perspective. *Medical Education* **28**(5): 394–399.

Houghton, G. & Wall, D. (2000). Trainers' evaluations of the West Midlands formative assessment package for GP registrar assessment. *Medical Teacher* **22**(4): 399–405.

Hu, L. & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* **6**(1): 1–31.

Jehu, D., Picton, C.J. & Cher, S. (1970). The use of notes in examinations. *British Journal of Educational Psychology* **40**: 353–357.

Marsh, H.W., Hau, K.T. & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes, dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling* **11**: 320–341.

Marton, F. & Saljo, R. (1976). On qualitative differences in learning: 1. Outcome and process. *British Journal of Educational Psychology* **46**: 4–11.

Mennin, S.P. & Kalishman, S. (1998). Student assessment. *Academic Medicine* **73**(9): S46–S54.

Michaels, S. & Kieran, T.R. (1973). An investigation of open-book and closed-book examinations in Mathematics. *The Alberta Journal of Educational Research* **19**(3): 202–207.

Paschal, C. (2002). Formative assessment in Physiology teaching using a wireless classroom communication system. *Advances in Physiology Education* **26**(4): 299–308.

Peat, M. & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology* **33**(5): 515–523.

Peitzman, S.J., Nieman, L.Z. & Gracely, E.J. (1990). Comparison of "fact recall" and "higher order" questions in multiple choice examination as predictors of clinical performance of medical students. *Academic Medicine* **65**(9 suppl): S59–S60.

Ramsden, P. (1992). *Learning to Teach in Higher Education* London: Routledge, Ch 5.

Ritter, L. (2000). The quest for an effective form of assessment: The evolution and evaluation of a Controlled Assessment Procedure (CAP). *Assessment & Evaluation in Higher Education* **25**(4): 307–320.

Rolfe, I. & McPherson, J. (1995). Formative assessment: how am I doing? *The Lancet* **345**: 837–839.

Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**: 119–144.

Schuwirth, L.W.T. & Vleuten, C.P.M.van der (2003). ABC of learning and teaching in medicine: Written assessment. *British Medical Journal* **326**: 643–645.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examinations versus assignment essay. *Higher Education* **35**: 453–472.

Smith, S.N. (2005). Learning approaches: Examination type, discipline of study, and gender. *Educational Psychology* **25**(1): 43–53.

Tang, K.C.C. (1992). Perception of task demand, strategy attributions and student learning. *Research and Development in Higher Education* **15**: 474–481.

Thissen-Roe, A., Hunt, E. & Minstrell, J. (2004). The DIAGNOSER project: Combining assessment and learning. *Behavior Research Methods, Instruments & Computers* **36**(2): 234–240.

Vaz, M., Avadhany, S.T. & Rao, B.S. (1996). Student perspectives on the role of formative assessment in physiology. *Medical Teacher* **18**(4): 324–326.

Velan, G., Kumar, R., Dziegielewski, M. & Walkefield, D. (2003). Web-based assessments in pathology with Questionmark Perception. *Pathology* **34**(3): 282–284.

Wass, V., Vleuten, C. van der, Shatzer, J. & Jones, R. (2001). Assessment of clinical competence. *The Lancet* **357**(9260): 945–949.

Weber, L.J., McBee, K. & Krebs, J.E. (1983). Take home tests: an experimental study. *Research in Higher Education* **18**(2): 473–483.