Advances in Health Sciences Education **9:** 15–27, 2004.

A Controlled Trial of Active Versus Passive Learning Strategies in a Large Group Setting

PAUL HAIDET[1]*, ROBERT O. MORGAN[1], KIMBERLY O'MALLEY[1], BETTY JEANNE MORAN[2] and BOYD F. RICHARDS[2]
[1]*Houston Center for Quality of Care and Utilization Studies and the Department of Medicine;*
[2]*Office of Curriculum, Baylor College of Medicine, Houston, Texas, USA*
*(*author for correspondence, Houston Veterans Affairs Medical Center, 2002 Holcombe Boulevard (152), Houston, Texas, 77030, USA; e-mail: phaidet@bcm.tmc.edu)*

Abstract. Objective: To compare the effects of active and didactic teaching strategies on learning- and process-oriented outcomes.
Design: Controlled trial.
Setting: After-hours residents'teaching session.
Participants: Family and Community Medicine, Internal Medicine, and Pediatrics residents at two academic medical institutions.
Interventions: We randomly assigned residents to two groups. One group received a didactic lecture on effective use of diagnostic tests; during this session, the teacher spent a full hour delivering content. The other group received the same content in a session structured to foster resident-to-resident interactions. In the latter session, the teacher spent only 30 minutes directly delivering content to residents.
Measures and Main Results: We measured residents' knowledge about and attitudes toward the session content before, immediately after, and one month after each session. We measured residents' perceptions of engagement and session value immediately after each session. We employed blinded observers who used a structured instrument to observe residents' activities during each session. Both teaching methods led to improvements in residents' scores on both knowledge and attitude assessments. The amount of improvement was not statistically different between groups. Residents in the active learning session perceived themselves, and were observed to be, more engaged with the session content and each other than residents in the didactic session. Residents in the didactic session perceived greater educational value from the session compared to residents in the active session. Conclusions: We reduced the amount of time spent in teacher-driven content delivery by 50 percent and covered the same amount of content with no detrimental effects on knowledge acquisition or attitude enhancement. Teaching strategies that foster learner-to-learner interactions will lead to more active engagement among learners, however, these learners may value the session less. Further research is needed to explore learner perceptions of the teaching process and other outcomes of active learning in medical education.

Key words: diagnostic tests – routine, education – medical, evidence-based medicine, internship and residency, learning, research

Introduction

There is an ongoing debate among medical educators about the most appropriate teaching method to employ in large-group settings. Proponents of didactic lectures argue that, given the large amounts of material that medical educators must cover during individual sessions, lectures allow the teacher to efficiently transfer content to learners (Brookfield, 1997; Bligh, 2000). On the other hand, many educators over the last fifteen years have pointed toward the need for teaching strategies that more actively engage learners with content and each other. It is thought that active learning methods lead to better conceptual structures and foster skills needed for life-long learning (Sutherland and Bonwell, 1996). Many interesting and innovative teaching methods that introduce active learning activities into large group sessions have been proposed (Elnicki et al., 1999; Green, 2000; Green and Ellis, 1997; Reilly and Lemon, 1997). We have been field-testing one such method, termed 'Team Learning' (Haidet et al., 2002; Seidel and Richards, 2001).

While both educational theory and common sense argue that active learning strategies should lead to improved learning outcomes over traditional didactic lectures, experimental data directly comparing active to didactic teaching strategies using the same content and learners are scarce. This absence of data exists at a time when leaders in medical education research have called for rigorous evidence in the selection of best teaching approaches (Van Der Vleuten et al., 2000; Griffith, 2000). Unfortunately, such evidence is difficult to gather. Medical education researchers encounter a variety of hurdles in designing comparative studies of teaching methods, including difficulty in measuring outcomes other than immediate knowledge acquisition, student performance bias (i.e., highly motivated students will learn content no matter what teaching method is used), and ethical barriers in randomizing students to different methods of teaching in the setting of established curricula (Cate, 2001; Bordage et al., 1998).

We performed this study to compare outcomes of an active learning teaching method to a traditional didactic lecture. We employed an experimental design that included an optional 'CME-type' setting for study participants and a variety of outcome measures that examined learner attitudes and the learning process in addition to traditional measures of knowledge acquisition.

Methods

Study design and recruitment

We invited all residents and chief residents from six different residency programs at two institutions (Family and Community Medicine, Internal Medicine, and Pediatrics at Baylor College of Medicine and the University of Texas, Houston College of Medicine) to attend a one-time session on 'effective use of diagnostic tests' and randomly assigned them to either a 'didactic' or 'active' session (described below). We allowed residents who were randomly assigned to a particular session

to cross over to the other session if their on-call duties precluded participation in their assigned session. In addition, we allowed residents who had not signed up prior to sign up as 'walk-ins' at the door at each study session. We offered residents a small financial incentive for participation, and we advertised that the content taught during the study sessions would have direct relevance to the Family and Community Medicine, Internal Medicine, and Pediatrics board certification exams. We informed residents that they were participating in a study to evaluate various teaching methods, but did not inform them whether their assigned session was the didactic or active session. We received approval for our recruitment and study protocol from the Baylor College of Medicine Institutional Review Board and obtained informed consent from all residents at study entry.

DESCRIPTION OF STUDY SESSIONS

We conducted the didactic and active sessions on successive evenings in a typical lecture hall on the Baylor College of Medicine campus. On both evenings, the same individuals interacted with the residents; these included a 'master of ceremonies' (BR), a teacher (PH), and administrative assistants. The 'master of ceremonies' introduced the teacher, kept time, and introduced all questionnaires using the same script for both sessions. The teacher conducted the educational portion of both sessions; this teacher had previous experience teaching the session content both through didactic and active methods in large group settings. For the didactic session, the teacher spent a total of 60 minutes delivering content using a standard 'slide and lecture' teaching method. For the active session, the teacher employed a 'Team-Learning' teaching method, described in detail previously (Haidet et al., 2002). Briefly, during the session he assigned learners to groups of 4–5 and gave each of these small groups a series of 3 tasks to be solved. In between tasks, the entire class engaged in a teacher-facilitated discussion of concepts and issues generated by the small-group task-solving process. The teacher directly delivered content to learners only during the discussions among the entire class (approximately 30 minutes); for the purposes of this study, during small-group activities the teacher did not interact with the learners.

The content delivered during both sessions consisted of concepts centering on the effective use of diagnostic tests (Sackett et al., 1991). These included conceptual and mathematical definitions of sensitivity, specificity, predictive value, and prevalence, and the use of these concepts in ordering and interpreting the results of diagnostic tests. In order to ensure uniformity of content among didactic and active sessions, we used content from the didactic session slides to construct the small-group activities for the active session. In addition, a 'readiness handout' containing the session goals and pertinent take-home points (identical for both sessions) was given to learners at the beginning of each of the sessions. This readiness handout appears in the Appendix.

MEASUREMENT OF BASELINE CHARACTERISTICS AND OUTCOMES

We used multiple measures to assess learning and process outcomes before, during, and after each session. As described in detail below, learning outcomes that we measured included assessments of knowledge and learner attitudes toward the usefulness and importance of session content. Process outcomes that we measured included learner patterns of interactions and perceptions of: (1) active engagement during the session; (2) the session's success in achieving stated learning goals; and (3) the session's future value. In addition to learning and process outcomes, we measured baseline characteristics of learners including age, year of residency, specialty, gender, residency institution, medical school graduated from (U.S. versus non-U.S.), advanced degrees, previous training in teaching or educational techniques, and ethnicity.

We assessed learning outcomes by measuring learner knowledge and attitudes before, immediately after, and one-month after each session. To measure learner knowledge, we invited nationally recognized experts in evidence-based medicine to write true-false and multiple choice questions related to the session goals. We asked each expert to write a variety of questions along a continuum of difficulty (i.e., 15% of learners should get the correct answer to 75%). We selected nineteen questions for our knowledge assessment from this pool based primarily on their relevance to the session goals and secondarily on their difficulty. We tested the same concepts at the same presumed level of difficulty on all three knowledge assessments (before, immediately after, and 1-month after). We varied the wording of the questions on these assessments so that they had different correct answers across the three time periods. We allowed learners the same length of time at both sessions to complete the pre- and immediate post-session knowledge assessments. The one-month knowledge assessment was mailed to study participants with instructions to complete it individually as a closed-book exam with no specific time limit. Prior to the sessions, we piloted the knowledge assessments for readability and clarity among general internal medicine fellows at the Houston Veterans Affairs Medical Center.

In addition to knowledge assessments, we measured learner attitudes toward session content before, immediately after, and 1-month after each session. We used two previously piloted scales (6 items for sensitivity/specificity, 5 items for predictive value) that measured learners' attitudes toward the usefulness of session concepts in their daily residency duties (Haidet et al., 2002). Examples of items from the sensitivity/specificity scale include: 'During my daily patient care duties, I think about the concepts of sensitivity and specificity'; and 'I always consider sensitivity/specificity when ordering diagnostic tests.' These scales demonstrated very good reliability (Cronbach alpha > 0.75 for both scales).

We measured learning process outcomes through a combination of assessments by neutral observers and learners themselves. We employed two trained observers blinded to the study purpose to observe learner patterns of interactions during each

session. The observers used a modification of the Stallings Observation Instrument (Stallings and Needles, 1985). High interobserver agreement using this instrument has been demonstrated in other educational settings (Swank et al., 1989). Briefly, this instrument required observers to choose specific learners and classify their actions at a single point in time. The observers chose learners at regular intervals throughout the session using a random selection scheme. Examples of learner actions recorded by the observers included: 'talking to the instructor', 'talking to other learners', 'listening to the instructor', and 'listening to other learners.'

To measure the residents' perceptions of the learning process, we convened a panel of education experts to develop instrument items that would measure learners' perceptions of their own engagement in the learning process and the future value of the sessions. We organized the items generated by this panel into two scales. The first scale measured learners' perceptions of engagement and contained 12 items including: 'I paid attention most of the time in class'; 'Most students were actively involved in class today'; and 'I was mostly a passive learner in class today (reverse scored).' The second scale measured learners' perceptions of future value of the sessions and contained 3 items including: 'In the future, I will be able to make appropriate use of what I learned in class today'; and 'I doubt that in the future I will remember most of the material from class today (reverse scored).' Items on both of these scales used 5-point Likert responses. We asked learners to complete both scales immediately following the session. We calculated an 'engagement' score and a 'value of the session' score as the mean of the items on each scale, respectively. Confirmatory factor analysis provided evidence for the unidimensionality of the items on each of these scales. Both scales demonstrated very good reliability (Cronbach alpha = 0.80 for the engagement scale and 0.75 for the value scale).

In addition to the assessments of engagement and value, we asked learners to rate their session with respect to each of the 4 stated objectives (see Appendix) on a 5-point scale anchored by 'Session met this objective extremely well' and 'Session did not meet this objective.' We calculated an overall score for how well the session met the objectives by calculating a sum of the scores on these four items.

STATISTICAL ANALYSIS

We used Student's t-test and chi-square analysis to compare continuous and categorical variables between groups, respectively. We used repeated measures analysis of variance to compare changes in knowledge and attitudes between the didactic and active groups from before to immediately- and one-month after the sessions. We report results as comparisons of means for continuous variables and comparisons of percentages for categorical variables.

Table I. Characteristics of study groups*

Characteristic	Didactic group	Active group
Age (mean ± sd)	28.6 ± 2.8	28.1 ± 2.9
First-Year Residents (%)	33	26
Internal Medicine Residents (%)	56	51
Female Gender (%)	52	42
Baylor Residents (%)	74	89
Graduated from Medical School Outside U.S.A. (%)	18	17
Advanced Degrees (MPH, PhD, etc) (%)	30	36
Advanced Training in Education (%)	18	11
Caucasian Ethnicity (%)	40	55

*$p > 0.05$ for all comparisons between the didactic and active groups.

Results

STUDY POPULATION

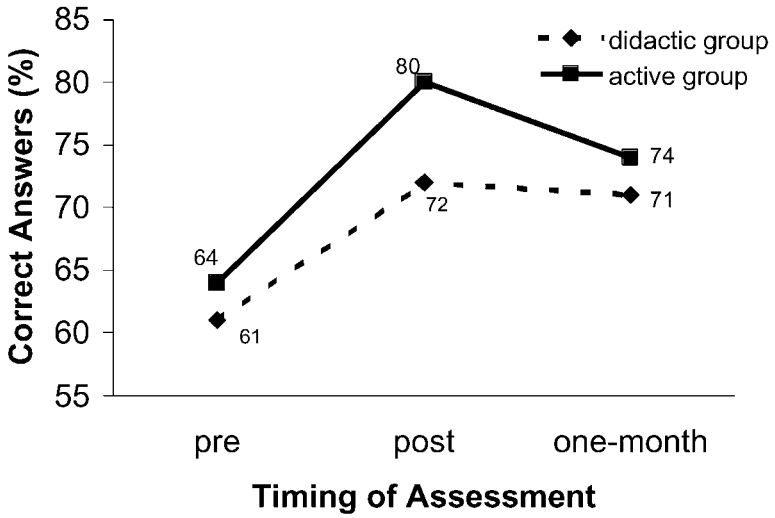
We enrolled and randomized 82 residents; 38 residents were assigned to the didactic group, and 44 were assigned to the active group. Of these, 24 residents 'crossed over'; 12 switched from the didactic to the active session, and 12 switched from the active to the didactic session. 12 residents did not show up for the didactic session, and 14 residents did not show up for the active session. There was one 'walk-in' for the didactic session and 6 'walk-ins' for the active session. This left a total of 27 learners in the didactic session and 36 learners in the active session.

Table I details demographic characteristics of both groups. There were no statistically significant differences between groups among any of the demographic characteristics we measured.

LEARNING OUTCOMES

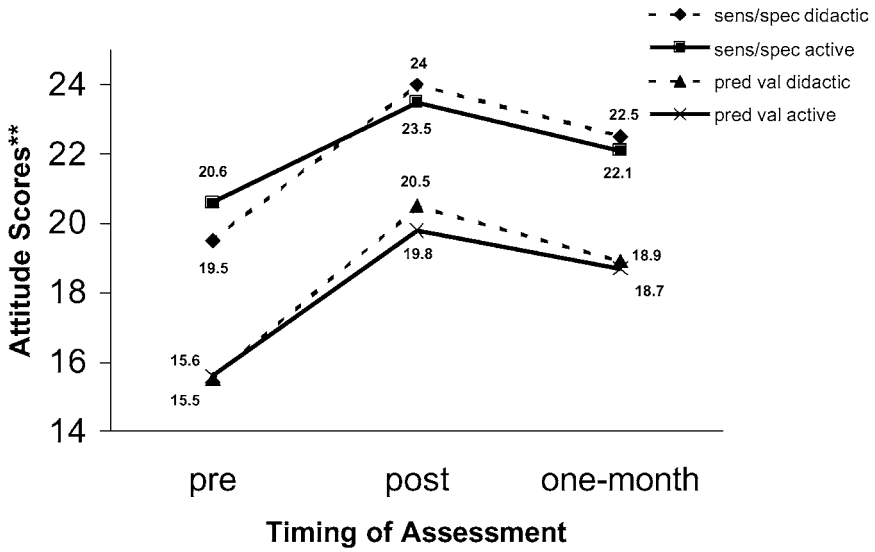
Results of the knowledge assessment are detailed in Figure 1. Both groups demonstrated significant gains and retention of knowledge ($p \leq 0.0003$ for effect of time period on overall scores). However, there were no significant differences between groups in either gain of knowledge or retention of knowledge. While eight participants (4 from the active group, 4 from the didactic group) did not complete the one-month followup, there were no significant differences on pretest and post-test scores between participants that completed followup and participants that did not.

Results of attitude assessments are detailed in Figure 2. Again, both groups demonstrated significant improvements in attitudes toward both sensitivity/specificity and predictive value content ($p < 0.0001$ for effect of time period on overall scores). However, the differences between groups in these attitude



*p value for effect of group assignment on change in scores = 0.67; N = 65 for pre and post test, N = 57 for one-month followup.

Figure 1. Knowledge assessments between study groups.*



*Sens/spec = sensitivity/specificity attitudes; pred val = predictive value attitudes.
p values for effect of group assignment on changes in attitudes over time are 0.22 for sens/spec attitudes and 0.68 for pred val attitudes.
**Higher scores indicate more favorable attitudes toward the content area assessed. Possible scores for sens/spec attitudes range from 6–30; possible scores for pred val attitudes range from 5–25.

Figure 2. Attitude assessments between study groups.

Table II. Patterns of interactions by learners

	Didactic group	Active group
Interacting with instructor	84%	57%
Interacting with student(s)	16%	43%

*$p < 0.001$, chi square analysis.

Table III. Learners' perceptions of the teaching process*

Scale	Didactic session mean score	Active session mean score	P value (student's t-test)
Engagement*	3.6	4.0	0.001
Value of the Session*	4.3	3.9	0.03
Met Objectives	6.1	9.5	0.004

*Higher numbers indicate more favorable responses (i.e., more actively engaged and higher perceived value of the session). The 'met objectives' scale is reverse scored; higher numbers indicate lower perceptions that the session met its objectives.

improvements were not significant for either sensitivity/specificity or predictive value attitudes.

PROCESS OUTCOMES

Patterns of learner interactions as measured by the trained observers appear in Table II. We were interested in patterns of interactions as a measure of student activity during the session. We defined 'interacting' as talking to or listening to another person or persons. As demonstrated in Table II, significantly more learners were interacting with each other during the active session, as opposed to mostly one-to-one interactions with the teacher during the didactic session. While the total number of 'listening' observations (regardless of who the learner was listening to) was similar in both groups (82 'listening' observations for the active group versus 71 observations for the didactic group), the total number of 'talking' observations was considerably less in the didactic group (22 'talking' observations for the active group versus 6 observations for the didactic group).

In addition to the observers' measurements, we solicited learners' perceptions of: (1) their engagement in the learning process; (2) educational value of the session; and (3) how well the session met its objectives. These results are detailed in Table III. While learners in the active session perceived themselves to be more actively engaged in the learning process, they perceived that the session had lower educational value and did not meet the objectives as well.

Discussion

This study provides evidence on the outcomes of an active-learning method compared to didactic lecturing in a large-group setting. The teacher was able to cover the same amount of conceptually complex and mathematically-oriented content in the active session as in the didactic session with no detrimental effects on short- or long-term knowledge acquisition or attitude enhancement. While the teacher in this study did not interact with the active learning participants during small group exercises for the purposes of the study design, teachers in real life settings would be able to interact with individual groups of learners during the small-group activities to further achieve their teaching aims. In addition to achieving the same knowledge and attitudes, learners in the active session also both self-rated and were observed to be more directly engaged with each other and the session content.

An interesting finding in this study is that the active session learners had lower perceptions of both the session's value and its ability to meet learning objectives even though they achieved the same knowledge and attitudinal gains. How can learners be more actively engaged but simultaneously perceive lower value? We propose several hypotheses for future study. Since the culture of medical education has traditionally emphasized the value and legitimacy of didactic lectures over other methods of instruction, we hypothesize that when learners found themselves in a situation where the methods were unusual (i.e., the small-group/large group format of the active session), these methods were marginalized and the session was seen as less useful. In a somewhat related vein, didactic teaching implies the need for an 'expert' to interpret and deliver complex content to novices (Palmer, 1998). In an active situation such as the small groups in this study, the learners were placed in a situation where they had to adapt to, interact with, and learn from other learners; while the actual learning of content was just as great as the didactic lecture, the reduction of contact time with the 'expert' may have lead to lower learner perceptions of the value of the session. Both of these hypotheses are driven by the premise that learners enter a teaching situation with a-priori assumptions about learning – assumptions that themselves have been learned during the acculturation process into the medical education environment (Becker et al., 1961). For medical educators who advocate greater use of active learning strategies, such assumptions need to be understood and addressed directly if the teaching is going to be seen to hold value by the learners.

An alternate hypothesis for the lower perceptions of value among learners in the active group has to do with the structure of the active learning session itself. During this session, the teacher quickly assigned learners (randomly) to small groups and immediately gave these groups a task to solve. The learners in the active group may have been reacting to their situation of having to work with strangers in intimate groups to achieve complex learning goals in a short amount of time. Educators have noted the importance of time, repeated contact, and familiarity among members of small groups (none of which were present in the active session)

in order for the group process to be effective (Michaelsen et al., 1997a). In addition, if small groups are to work to their maximum potential, there must be built into the learning environment incentives for small groups to work effectively together (Michaelsen et al., 1997b; Johnson et al., 2000). Our finding that the active group learners demonstrated comparable knowledge and attitude gains to the didactic group despite the absence of conditions to foster high quality group interactions suggests the possibility that active learning strategies may in fact lead to better learning outcomes when conditions are met that foster both active engagement *and* effective interpersonal interactions.

In the last several years, there have been repeated calls for evidence in medical education (Van Der Vleuten, 2000; Griffith, 2000; The Best Evidence Medical Education Group, 2000; Lynch et al., 2000; Wolf, 2000; Harden and Lilley, 2000; Davies, 2000). These calls have emphasized the need for rigorously conducted studies that employ experimental designs to examine causal relationships and educational effectiveness. In this study, we attempted to address such calls and would like to highlight two important methodological features. First is the setting in which we conducted the teaching sessions. Past studies in the setting of ongoing medical education have encountered student performance bias and ethical barriers to the employment of experimental design strategies. By conducting our study outside of the residencies' curricula, we were able to ensure confidentiality by removing performance measures from the purview of residency supervisors, and we were able to obtain informed consent and ethically apply a randomization scheme. In addition, since there were no grades or incentives to perform well on our learning assessments, we assume a reduced effect of student performance bias that has fostered regression to the mean in other studies of educational effectiveness (Cate, 2001).

A second methodological feature that we would like to highlight is our measurement of learning and process outcomes. Others have proposed that measurement of additional educational outcomes beyond short-term knowledge is a priority for medical education research (Bordage et al., 1998). In this study, we measured not only short-term knowledge, but also long-term outcomes, attitudinal outcomes, and process outcomes. In addition, we employed a combination of objective and perceptual measures to assess these outcomes. The use of multiple outcome measures in this study led not only to conclusions about the question of the relative efficacy of different learning approaches, but also to the illumination of the interesting tension between learners' performance and learners' perceptions of value that deserves further study. We plan future work to develop measures for other educational outcomes, such as effective team communication skills, that may be pertinent to effective clinical practice.

This study has several limitations. First, the large percentage of residents in each group who were no-shows or who self-selected themselves into their group compromises the randomization scheme. This limitation is tempered by the fact that the residents did not know a-priori what teaching methods would be used; self-

selection was therefore unlikely to be based on a particular teaching method or to affect the impact of the different methods. In addition, the final study groups did not differ in any of the potentially confounding characteristics that we measured. A second limitation is that the small numbers of learners in each group limits the power of this study to find statistically significant differences in knowledge and attitude outcomes between the groups. However, the absolute value of the effect sizes between groups in this study was exceedingly small, suggesting that no educationally meaningful differences actually exist.

In conclusion, we have studied various outcomes of didactic versus active teaching methods in a large-group setting and demonstrated that active learning strategies can be employed in the lecture hall without reductions in knowledge acquisition or attitude enhancement by learners. In addition, small-group task solving work within a large-group session will produce more active engagement among learners, but leads to lower learner perceptions of the session's value. Further educational research should be directed toward the factors associated with learner perceptions of teaching value and the effects of such perceptions on learning outcomes.

Acknowledgements

The authors would like to acknowledge Dovie Dupree, Kenda Moritz, Cindy Schott, and Tenecia Williams for assistance with subject recruitment, session administration, and data management. The authors would also like to thank all of the individuals who contributed questions for the knowledge assessment used in this study.

Supported by grant #P116B000884 from the U.S. Department of Education Fund for the Improvement of Post-Secondary Education (FIPSE), and with funds from the Master Teachers' Fellowship Program at Baylor College of Medicine, Houston, Texas. Dr Haidet is supported by a career development award from the office of Research and Development, Health Services R&D Service, U.S. Department of Veterans Affairs.

Appendix: Readiness Handout (Given to Learners at Beginning of Active and Didactic Sessions)

Learning objectives:

1. Define and understand the concepts of sensitivity and specificity.
2. Define and understand the concepts of positive and negative predictive value.
3. Understand the concept of prevalence and if/how it affects the above items.
4. Be able to have a feel for the usefulness of the information that a test gives in a particular situation.

Some definitions (in English):

Sensitivity – "out of all the people who have the disease, this is the proportion that test positive"
Specificity – "out of all the people who **don't** have the disease, this is the proportion who test negative"
Positive predictive value – "out of all the people who test positive, this is the proportion that actually have the disease"
Negative predictive value – " out of all the people who test negative, this is the proportion that actually **don't** have the disease"

Some definitions (not in English):

	Diseased	Not diseased
(+) Test	a	b
(–) Test	c	d

a+c = e

Prevalence = e/N N = Total Population Number (a+b+c+d)

Sensitivity = a/(a+c)

Specificity = d/(b+d)

Pos Pred Value = a/(a+b)

Neg Pred Value = d/(c+d)

References

Becker, H.S., Geer, B., Hughes, E.C. & Strauss, A.L. (1961). *Boys in White. Student Culture in Medical School* (pp. 80-187). New Brunswick: Transaction.

Bligh, D.A. (2000). *What's The Use of Lectures?* San Francisco: Jossey-Bass.

Bordage, G., Burack, J.H., Irby, D.M. & Stritter, F.T. (1998). Education in ambulatory settings: developing valid measures of educational outcomes and other research priorities. *Academic Medicine* **73**: 743–750.

Brookfield, S.D. (1990). Lecturing creatively. In *The Skillful Teacher* (pp. 71–87). San Francisco: Jossey-Bass.

Cate, O.T. (2001). What happens to the student? The neglected variable in education outcomes research. *Advances in Health Sciences Education* **6**: 81–88.

Davies, P. (2000). Approaches to evidence-based teaching. *Medical Teacher* **22**: 14–21.

Elnicki, D.M., Halperin, A.K., Shockcor, W.T. & Aronoff, S.C. (1999). Multidisciplinary evidence-based journal clubs: curriculum design and participants' reactions. *American Journal of Medical Sciences* **317**: 243–246.

Green, M.L. & Ellis, P.J. (1997). Impact of an evidence-based medicine curriculum based on adult learning theory. *Journal of General Internal Medicine* **12**: 742–750.

Green, M.L. (2000). Evidence-based medicine training in internal medicine residency programs: a national survey. *Journal of General Internal Medicine* **15**: 129–133.

Griffith, C.H. (2000). Evidence-based educational practice: the case for faculty development in teaching. *American Journal of Medicine* **109**: 749–751.

Haidet, P., O'Malley, K. & Richards, B. (2002). An initial experience with 'Team Learning' in medical education. *Academic Medicine* **77**: 40–44.

Harden, R.M. & Lilley, P.M. (2000). Best evidence medical education: the simple truth. *Medical Teacher* **22**: 117–119.

Johnson, D.W., Johnson, R.T. & Smith, K.A. (2000). Constructive controversy. *Change* **32**: 29–37.

Lynch, D.C., Whitley, T.W. & Willis, S.E. (2000). A rationale for using synthetic designs in medical education research. *Advances in Health Sciences Education* **5**: 93–103.

Michaelsen, L.K., Black, R.H. & Fink, L.D. (1997a). What every faculty developer needs to know about learning groups. In DeZure, D. (ed.), *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development*. Stillwater, Ok: New Forums Press.

Michaelsen, L.K., Fink, L.D. & Knight, A. (1997b). Designing effective group activities: lessons for classroom teaching and faculty development. In DeZure, D. (ed.), *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development*. Stillwater, Ok: New Forums Press.

Palmer, P. (1998). *The Courage to Teach. Exploring the Inner Landscape of a Teacher's Life* (p. 100). San Francisco: Jossey-Bass.

Reilly, B. & Lemon, M. (1997). Evidence-based morning report: a popular new format in a large teaching hospital. *American Journal of Medicine* **103**: 419–426.

Sackett, D.L., Haynes, R.B., Guyatt, G.H. & Tugwell, P. (1985). *Clinical Epidemiology: A Basic Science for Clinical Medicine*, second edition (pp. 77–94). Boston: Little, Brown, and Company.

Seidel, C.L. & Richards, B.F. (2001). Application of team learning in a medical physiology course. *Academic Medicine* **76**: 533–534.

Stallings, J. & Needles, M. (1985). *Stallings Observation Instrument* (rev. ed.). Stanford: SRI International.

Sutherland T.E. & Bonwell C.C. (eds.), (1996). *Using Active Learning in College Classes: A Range of Options for Faculty*. San Francisco: Jossey Bass.

Swank, P.R., Taylor, R.D., Brady, M.P. & Freiberg, H.J. (1989). Sensitivity of classroom observation systems: measuring teacher effectiveness. *Journal of Experimental Education* **57**: 171–86.

The Best Evidence Medical Education Group (BEME) (2000). Report of meeting – 3–5 December 1999, London, UK. *Medical Teacher* **22**: 242–245.

Van Der Vleuten, C.P.M., Dolmans, D. & Scherpbier, A. (2000). The need for evidence in education. *Medical Teacher* **22**: 246–250.

Wolf, F.M. (2000). Lessons to be learned from evidence-based medicine: practice and promise of evidence-based medicine and evidence-based education. *Medical Teacher* **22**: 251–259.