

Peter Spear, Ph.D.
Provost
University of Wisconsin
Madison, Wisconsin

August 21, 2002

Dear Dr. Spear,

Please find enclosed the assessment report for the school of medicine for the period 1996-2001. If you have any questions about the material contained in the report or would like additional information, please feel free to contact me.

Sincerely,

Mark Albanese, Ph.D.
Professor of Population Health
Director, Medical Education Research and Development

cc. Phillip Farrell, M.D., Ph.D.
Susan Skochelak, M.D., MPH
Eden Inoway, Ph.D.
Carolyn Bell, M.D.

Assessment Report for the University of Wisconsin Medical School (UWMS)
1996-2001

Mark Albanese, Ph.D., Director, Medical Education Research and Development, Professor of Preventive Medicine

Date: August 21, 2002

A comprehensive curriculum evaluation system was approved by the Educational Policy Committee and presented to the faculty-at-large at Medical Education Day 1994. This report describes each element of the system and provides a brief summary of our most recent results. The period covered is from 1996-01.

I. Evaluation System of the UW Medical School

The comprehensive evaluation system has the following features:

1. Formative evaluation support services are available to course directors and faculty to help them develop effective student, instructor and course evaluation mechanisms as well as interpret the results. The UWMS Testing and Evaluation Service provides expert consultation on test development methods and interpreting test analysis statistics. It also provides course directors with test support assistance which includes providing machine scannable answer sheets for tests, copying written tests, proctoring of exams, scoring of written exams, producing exam score reports for course directors, reporting scores to students using WebCT and similar assistance in administering tests via computer. The UWMS Testing and Evaluation Service also administers tests to comply with Americans with Disability Act accommodations. Appendix A contains an example of a test report provided by the UWMS Testing and Evaluation Service. For instructor and course evaluation, the UWMS Testing and Evaluation Service administers the forms to students at a time and in the method preferred by course directors, analyzes the results and provides a summary report. Appendix B contains an example of an instructor evaluation and course report provided by the UWMS Testing and Evaluation Service. Appendix C provides a table describing the formative evaluation services used by courses across the four years of the curriculum.

2. Formative evaluation data are collected and reported back to teaching faculty for the purposes of continuous quality improvement. The School of Medicine has adopted a common student evaluation form for use across the curriculum; additionally, however, a small number of courses still have their own forms that they administer. There are separate forms for use in each year of the curriculum as well as a form specifically used to evaluate examinations (Appendix D). The standard evaluations are administered through the Office of Medical Education Research and Development (OMERD) via the UWMS Testing and Evaluation Service so as to be independent of course directors. End of semester student focus groups are also conducted to determine to what extent the courses taken were effectively integrated and to obtain details that could help improve courses. The associate dean for curriculum, director of undergraduate programs and director of OMERD meet with course directors, individually, at the end of each semester to provide the formative evaluation data from student course evaluations and student focus groups to celebrate successes and discuss plans for addressing weaknesses. There are also monthly course director group meetings to discuss larger curriculum issues.

3. A longitudinal data-base profiling student performance across classes has been created to detect potential differences due to curriculum changes across time. The data base

extends back to 1986 in order to provide a sufficiently long perspective to interpret any changes that may be noted. This data base includes pre-medical GPAs and MCAT scores, grades in medical school, licensure examination performance data, and post-graduate data described in items 3 and 4 below.

4. Alumni evaluations of their medical education at post-graduate year (PGY) 1 are solicited. The PGY 1 evaluation forms are administered in August after graduates have completed their first year of residency. They are structured to give an assessment of their own level of medical preparation in comparison to their colleagues from other medical schools. For 1994-2000 graduates, the response rates and the percentage of graduates rating their medical education in the very good to outstanding range are shown in table 1.

Table 1

Response Rates and Percentages Rating their Undergraduate Medical Education Very Good to Excellent: 1994-2000 Graduates

Graduation Year	1994	1995	1996	1997	1998	1999	2000
Response Rate	67%	63%	57%	36%	53%	41%	61%
% Rated Very Good-Excellent*	93%	86%	95%	90%	92%	90%	88%
Graduation Year	2001						
Response Rate	43% **						
% Rated Very Good-Excellent*	90% **						

* Rating Scale: 1=poor, 2=fair, 3=good, 4=very good, 5=excellent, 6=outstanding

** Received as of September 15, 2002. Second reminder mailing in progress.

The response rate ranged from 36% to 67%. With the exceptions of unusually low response rates in 1997 and 1999 (2001 data are still being processed), we generally have had over 50% of the graduates respond to the survey. The percentage of graduates rating their undergraduate medical education very good to excellent ranged from 86% to 95% and seems to be hovering around 90% over the last five years.

6. Supervisor ratings of the performance and level of preparation of graduates at end of PGY-1 are also administered in August after graduates have completed their first year of residency. They are structured to give an assessment of our graduates' level of medical preparation in comparison to their colleagues from other medical schools. For our 1994-2000 graduates, the response rates and the percentage of supervisors rating our graduates' performance in the very good to outstanding range are shown in table 2.

Table 2

Response Rates and Percentages of Supervisors Rating UW Graduates Very Good to Excellent: 1994-2000 Graduates

Graduation Year	1994	1995	1996	1997	1998	1999	2000
Response Rate	90%	95%	80%	70%	64%	79%	88%
% Rated Very Good-Excellent*	73%	81%	89%	86%	76%	88%	78%
Graduation Year	2001	2002	2003	2004	2005	2006	2007
Response Rate	61% **						
% Rated Very Good-Excellent*	83% **						

* Rating Scale: 1=poor, 2=fair, 3=good, 4=very good, 5=excellent, 6=outstanding

** Received as of September 15, 2002. Second reminder mailing in progress.

The response rate ranged from 64% to 95 (2001 results are still in progress). The percentage of supervisors rating UW graduates very good to excellent ranged from 73% to 89%. While the trend from 1994 to 1996 was progressively more positive, the results since 1997 have ranged from 76%-88% in a relatively random pattern. Generally, these data suggest that our graduates are performing well in residency.

7. The year-end professional skills assessment (YEPSA), an objective structured clinical examination, is administered at the end of the third year of the medical curriculum to ensure students have the basic clinical skills they need to proceed with their medical education. Information from this examination is analyzed and given back to course directors as feedback and presented to the Educational Policy Committee to determine if the curriculum is meeting the needs of our students. In the first operational implementation in 1997, all students passed the overall exam (passing was set at 70%), but there were about 10 who failed individual stations. Students failing individual stations were given personalized remediation assignments. The results from 1998 found 8 students failing at the 70% criterion level and another 38 students needing remediation on individual stations.

In 1999, a task force set about to review the examination and establish agreed upon passing standards. The revised passing standard was based upon overall ratings of performance instead of accumulation of points. For the 10-12 station examination, students receiving fail ratings on 2 or more stations, failed the examination. Ratings of marginal and below were considered to indicate that students needed remediation on a station and for passing purposes, 2 marginal ratings equaled one failure. Thus, a student receiving a marginal rating on four or more stations would fail the overall examination. At the same time the new rating system was implemented, passing YEPSA was made a graduation requirement. Based upon the changes implemented in 2000, we had 6 failures of the overall examination and 84 students who need remediation on one or more stations. In 2001, we had almost identical results with 6 failures of the overall exam and 83 students who needed remediation on one or more stations. In 2002, there were 7 failures overall and 77 students needing remediation on one or more stations. These data suggest that the procedures implemented in 2000 have produced relatively stable results. Table 3 summarizes the YEPSA data from 1998 to date.

Table 3

YEPSA Results for 1998-2002 Graduates

Graduation Year	1998	1999	2000	2001	2002
Number of Overall Failures	0	6	6	6	6
Number Needing Station Remediation	10	38	84	83	77
Number Taking YEPSA	135	148	143	125	138

8. Licensure Examination Results

The National Board of Medical Examiners administers the United States Medical Licensure Examination (USMLE) whose passing is required for all physicians who wish to practice medicine in the United States. The USMLE consists of three tests or steps. Step 1 is usually taken by students between the second and third year of medical school. Step 2 is usually taken in the fourth year of medical school. Step 3 is taken in the second year of residency (Post graduate year 2, PGY2). We monitor performance of our students on Steps 1 and 2. We have data on the percentage of our graduates who fail each of the steps as well as means and standard deviations for our students as well as the mean expressed in effect size units (ES), where the ES is computed as follows:

$$ES = (\text{UW Mean} - \text{National Mean}) / \text{National Standard Deviation}$$

Table 4 shows UW Medical School data for graduating classes of 1988 through 2003 (Step 1) and 1988 through 2001 (Step 2).

Table 4

USMLE Step 1 and Step 2 Data for UW Medical School Students of Graduating Classes
1989-2003

Step 1					Step 2				
Year Graduated	Year Step 1 Was Taken	% Fail	Mean	Effect Size	Year Graduated	Year Step 2 Was Taken	% Fail	Mean	Effect Size
1988	1985-86	**	**	**	1988	1987-88	0	560	0.57
1989	1986-87	8	519	0.23	1989	1988-89	0	543	0.46
1990	1987-88	14	494	0.13	1990	1989-90	0	512	0.21
1991	1988-89	7	513	0.19	1991*	1990-91	2	204	0.20
1992	1989-90	7	509	0.24	1992	1991-92	2	205	0.29

1993*	1990-91	8	204	0.20	1993	1992-93	1	209	0.43
1994	1991-92	6	206	0.29	1994	1993-94	5	207	0.30
1995	1992-93	3	208	0.25	1995	1994-95	5	203	0.22
1996	1993-94	5	209	0.23	1996	1995-96	5	209	0.30
1997	1994-95	6	207	0.00	1997	1996-97	1	210	0.13
1998	1995-96	3	212	0.09	1998	1997-98	2	216	0.30
1999	1996-97	9	209	-0.14	1999	1998-99	3	218	0.35
2000	1997-98	5	217	0.10	2000***	1999-00	5	221	0.33
2001	1998-99	3	215	0.00	2001	2000-01	4	216	0.04
2002***	1999-00	5	215	0.00	2002	2001-02	4	222	
2003	2000-01	8	217	.08					
2004	2001-02	7	218						
2005	2002-03								

*Scoring scale was changed in 1991 (From a mean of ~500 to ~200)

**Data not available at this time

***Step 1 Passing score increased from 176 to 179, Step 2 increased from 179-182

The USMLE data are relatively promising, especially those for Step 2. The Step 2 results ranged from .1-.5 standard deviations above the national mean, and have been stable at .30-.35 standard deviations above the national mean for the past three years. While the failure rate on Step 2 rose in 2000 from 3% to 5%, the rise coincided with an increase in the passing score from 179 to 182. Thus, it is most likely an artifact of the change in the passing score.

The Step 1 results are somewhat of concern since 1997. For the five years prior to 1997, the Step 1 means were above the national mean by from .2-.3 standard deviations. Since that time, they have hovered around the national mean within +/- .2 standard deviations.

The data from the evaluation system has been used for providing feedback to individual course directors, presenting information for policy decisions to faculty committees and to provide information in deliberations of task forces on curriculum flexibility, admissions, and grading.

II. Assessment Plans for 2002-3

We plan to continue the activities described above, updating all information as new information becomes available. We are also reviewing all student surveys we administer to determine if we can reduce the number and still meet our information needs. In addition, we are reviewing all data collection procedures so as to reduce redundancy and determine if we can process data more efficiently and quickly. A major initiative is to determine what IRB protocols are necessary in order to make our assessment system as useful as possible. Given that the medical school desires to raise its visibility and national ranking, there is substantial interest in contributing to the enduring literature on how to educate medical students. This will inevitably involve using data from our evaluation system for publication. It is our goal to make sure that all data is collected in a form that is of highest validity and yet can serve this purpose.

We continue to be interested in applying statistical process control procedures to the management of course evaluation data as well as exploring its application to student academic performance data. The goal is to obtain early warning of problems to enable fixes to be implemented before small problems become large problems. Developing the infrastructure for this initiative has proven to exceed our available resources. This is likely to need grant support in order to proceed to the pilot testing stage.

The UWMS is developing competencies for graduating students with the intent of implementing them when we move to the new Health Sciences Learning Center in 2004. We are investigating use of Personal Data Assistants (PDA) for maintaining records of accomplishing competencies. We are also investigating electronic portfolios as a mechanism for building records of documentation of accomplishments.

III. Budgetary Needs for 2002-3

Approximately one-fourth of OMERD personnel time is directed at these activities and 75% of our supplies and expenses. In 2001-2, this amounted to \$128,829 in salaries and fringe benefits and \$19,506* in supplies and expenses. In 2002-3 our resource needs are likely to be greater. While we had hoped to make progress in our investigations into applications of statistical process control (SPC), we did not have the resources necessary to accomplish this goal. Instead, we explored computer administration of exams and on-line grading.

IV. Changes Made as a Result of Assessment Activities

We have provided increased support for some course directors based upon the results of some of our assessment activities (e.g., funding for teaching assistants). In other cases, assessment activities have led to changes in course directors; thus far, the changes in course directors have been at the instigation of the course director him/herself.

V. An Outstanding Experiment or Interesting Result Obtained Involving Assessment

We have continued to get increasing access to the data ware-house on the people soft system. This will enable us to access undergraduate grades for students who matriculate to medical school. This will be very helpful for evaluating our Medical Scholars Program. We continue to explore moving our tests and evaluations to on-line administration via WebCT. A lack of adequate facilities proximate to the medical school to administer computer examinations to our students has hampered progress in this area. We are investigating options around campus that might better meet our needs. In addition to the logistics of administering tests by computer, we face the challenge of introducing UWMS faculty to the variety of functions available in WebCT and helping faculty to use WebCT in their courses.

VI. Assessments That Did Not Work

Our attempts to move to testing and student evaluations on WebCT were only modestly successful. The attempt for students evaluations was particularly problematic.

c:\omerd\prov-rep02.doc

n:\acad\rd\albanese\comerd\prov-rep02.doc