Interactive-engagement methods in introductory mechanics courses*

Richard R. Hake[a)]
Department of Physics, Indiana University, Bloomington, Indiana 47405

A previous report [R.R. Hake, Am. J. Phys. **66**, 64-74 (1998)] of mechanics test data for 62 introductory physics courses with total enrollment of 6542 students strongly suggested that classroom use of interactive engagement (IE) methods can increase mechanics-course effectiveness in both conceptual understanding and problem-solving well beyond that achieved by traditional methods. This article is intended to assist (a) instructors in selecting and implementing IE methods, and (b) physics-education researchers in assessing and utilizing the raw data of the survey. Test scores, instuctional methods, materials used, institutions, and instructors for each of the survey courses are tabulated and referenced. Suggestions for the mitigation of various implementation problems are given, based on case studies of seven atypical courses which employed IE methods but achieved low normalized gains characteristic of traditional methods. Some research questions raised by the present survey and amenable to experimental investigation are posed.

I. INTRODUCTION

In order to try to gauge the effectiveness of various current introductory-mechanics-course educational methods, I initiated a survey of pre/post test results in 1992. Use was made of the well-known Halloun-Hestenes Mechanics Diagnostic[1](MD) or more recent Force Concept Inventory[2a,b] (FCI) tests of conceptual understanding, and the Hestenes-Wells Mechanics Baseline[3] (MB) test of problem-solving ability. Preliminary results[4a,b] were followed by abbreviated summary reports[5] which strongly suggested that ***classroom use of interactive engagement (IE) methods can increase mechanics course effectiveness in both conceptual understanding and problem-solving well beyond that achieved with traditional (T) methods.*** As discussed in ref. 5a, this conclusion is not abrogated by the fact that the method of data solicitation had a built-in bias towards relatively effective IE courses.

For survey classification and analysis purposes I defined[5a]:

(a) "Interactive Engagement" (IE) methods as those *designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors,* all as judged by their literature descriptions;

(b) "Traditional" (T) courses as those reported by instructors to *make little or no use of IE methods, relying primarily on passive-student lectures, recipe labs, and algorithmic-problem exams;*

(c) "Interactive Engagement" (IE) courses as those reported by instructors to *make substantial use of IE methods*;

(d) average normalized gain <g> for a course as the ratio of the actual average gain <G> to the maximum possible average gain, i.e.,

$$\langle g\rangle \equiv \%\langle G\rangle \;/\; \%\langle G\rangle_{max}$$
$$= (\%\langle S_f\rangle - \%\langle S_i\rangle)/(100 - \%\langle S_i\rangle), \quad \text{..............................(1)}$$

where $\langle S_f\rangle$ and $\langle S_i\rangle$ are the final (post) and initial (pre) class averages;

---

*Submitted on 6/19/98 to the potential new *Journal of Physics Education Research.* Comments and criticisms will be welcomed at: R.R. Hake; 24245 Hatteras St., Woodland Hills, CA 91367, USA, <hake@ix.netcom.com>.
A few very minor corrections and additions were made on 6/27/98.

(e) "High-g" courses as those with $(\langle g\rangle) \geq 0.7$;
(f) "Medium-g" courses as those with $0.7 > (\langle g\rangle) \geq 0.3$;
(g) "Low-g" courses as those with $(\langle g\rangle) < 0.3$.

In ref. 5a, a consistent analysis over diverse student populations with widely varying initial knowledge states, as gauged by $\langle S_i\rangle$, was obtained by taking the normalized average gain $\langle g\rangle$ as a rough measure of the effectiveness of a course in promoting conceptual understanding:

(a) *All* points for the 14 T courses (N = 2084) fell in the Low–g region, with

$$\langle\langle g\rangle\rangle_{14T} = 0.23 \pm 0.04\text{sd}. \quad \text{...} \quad (2a)$$

Here and below, double carets "$\langle\langle X\rangle\rangle_{NP}$" indicate an average of averages, i.e., an average of $\langle X\rangle$ over N courses of type P, and sd $\equiv$ standard deviation (*not* to be confused with random or systematic experimental error [5a]).

(b) Eighty-five percent (41 courses, N = 3741) of the 48 IE courses fell in the Medium-g region and 15% (7 courses, N =717) in the Low-g region. Overall,

$$\langle\langle g\rangle\rangle_{48IE} = 0.48 \pm 0.14\text{sd}. \quad \text{...} \quad (2b)$$

(c) No course points lay in the "High-g" region.

The interactive engagement courses were, on average, more than twice as effective as traditional courses in promoting conceptual understanding since $\langle\langle g\rangle\rangle_{IE} = 2.1\ \langle\langle g\rangle\rangle_T$. The difference $\langle\langle g\rangle\rangle_{48IE} - \langle\langle g\rangle\rangle_{14T} = 0.25$ is 1.8 standard deviations of $\langle\langle g\rangle\rangle_{48IE}$ and 6.2 standard deviations of $\langle\langle g\rangle\rangle_{14T}$, reminiscent of that seen in comparing instruction delivered to students in large groups with one-on-one instruction, as discussed in ref. 5a. It is extremely unlikely[5a] that systematic error played a significant role in the large difference observed in the average normalized gains of the T and IE courses.

The present article is intended to assist (a) instructors in selecting and implementing proven IE methods and (b) physics-education researchers in assessing and utilizing the raw data of the survey. I tabulate, discuss, and reference the particular methods and materials that were employed in each of the 62 survey courses. Suggestions for the avoidance of various implementation problems are given, based on case studies of seven atypical courses which employed IE methods but achieved low normalized gains characteristic of traditional methods. The present information, largely omitted from the abbreviated summary reports,[5] allows answers to three questions of interest to physics instructors and physics-education researchers:

Q1. *What methods and materials are being used in IE courses; where are descriptions and materials available; what are the types of institutions, characteristics of the students, and educational contributions of the instructors?*

Q2. *Are there any pedagogical difficulties in implementing IE methods, and if so, how might these be mitigated*?

Q3. *Does the present study give rise to any research questions calling for further experimental investigation?*

II. RAW DATA

The test data in Table I and the corresponding instructional methods in Table II were obtained from published accounts or private communications (see references for Tables I, II). [For presentation of these data in the form of gain (posttest – pretest) vs pretest graphs see ref. 5a.] The private communications usually included responses to a survey form[4c] which asked for information on the pre/post testing method; statistical results; institution; type of students; activities of the students; and the instructor's educational experience, outlook, beliefs, orientation, resources, and teaching methods. Aside from its survey purpose, the form's list of physics-education strategies and resources may be useful.

A. Pre/post Test Data

Tables Ia,b,c contain pre/post test data for 62 introductory courses enrolling a total of 6542 students using the conceptual Mechanics Diagnostic[1a] (MD) or Force Concept Inventory[2a,b] (FCI) exams, and (where available) the problem-solving Mechanics Baseline[3] (MB) test, always given as a posttest. The **bold-faced** data indicate an average normalized gain ($<g>$) ≥ 0.6 (only 12 of the survey courses - discussed in ref. 5a - achieved such gains). The instructors' names are given in the references (column 10), and instructors' initials are sometimes indicated in the "Course Code" (column 1). The courses listed are of two types (column 4): (1) "Traditional" (T) courses and (2) "Interactive Engagement" (IE) courses, both as defined in the Introduction.

To increase the statistical reliability[5a] of averages *over courses*, only those with enrollments $N \geq 20$ are listed in Tables I and II, although in some cases of fairly homogeneous instruction and student population (AZ-AP, AZ-Reg, WP92-C, TO, TO-C) courses or sections with less than 20 students were included in a number-of-student weighted average. In assessing the FCI, MD, and MB scores it should be kept in mind that the random guessing score for each of these five-alternative multiple-choice tests is 20%. However, completely non-Newtonian thinkers (if they can at the same time read and comprehend the questions) may tend to score *below* the random guessing level because of the very powerful interview-generated distractors which include most of the common mechanics misconceptions.[1,2a]

Table Ia. Pre/post test data for 14 introductory *high-school* physics courses enrolling a total of N = 1113 students. The **bold-faced** data indicate an average normalized gain (<g>) ≥ 0.6. Footnotes are placed after Table Ic.

COURSE CODE [a] [1]	LEVEL [b] [2]	N [c] [3]	TYPE [d] [4]	PRE [e] [StdD] % [5]	POST [e] [StdD] % [6]	GAIN [f] G % [7]	NORM. GAIN [g] g [8]	MB [h] [StdD] % [9]	REF. [10]
AZ-AP (3 cour.)	HS	33	T	41 (16)	57 (18)	16	0.27	39 (15)	2a
AZ-Hon (2 cour.)**	HS	62	T	27	45	18	0.25	—	2a
AZ-Reg (18 cour.)	HS	612	T	27 (11)	48 (16)	21	0.29	32 (11)	2a
BC-Hon	**HS**	**22***	**IE**	**32**	**79 (14)**	**47**	**0.69**	**52 (13)**	**6a**
BC-Reg	HS	43*	IE	24	50 (13)	26	0.34	—	6b
Chicago-Reg	HS	56	T	27	42	15	0.21	—	2a
CL-Reg	HS	20*	IE	35	62	27	0.42	—	7a
ELM-Hon	**HS**	**20***	**IE**	**18**	**74**	**56**	**0.68**	**—**	**8**
GS-Hon	HS	63	IE	28	66	38	0.53	47	2a
GS-Hon95	**HS**	**49**	**IE**	**28**	**72**	**44**	**0.61**	**56**	**2c**
LT-Hon	HS	27*	IE	30 (12)	70 (18)	40	0.57	—	9a
MW-Hon	**HS**	**30**	**IE**	**42 (18)**	**78 (15)**	**36**	**0.62**	**62 (17)**	**2a**
RM94-Reg	HS	38*	IE	33 (15)	65 (19)	32	0.48	—	10a
RM95-Reg	HS	38*	IE	32 (13)	67 (17)	35	0.51	—	10a

Table Ib. Pre/post test data for 16 introductory *college* physics courses enrolling a total of N = 597 students. Please refer to the heading of Table Ia for important explanations.

COURSE	LEV-	N [c]	TYPE [d]	PRE [e]	POST [e]	GAIN [f]	NORM.	MB [h]	
CODE [a]	EL [b]			[StdD]	[StdD]	G	GAIN [g]	[StdD]	REF.
				%	%	%	g	%	
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
DB-C	C4	27*	IE	36	59	23	0.36	—	11
LF92	C2	25	T	36	49	13	0.20	—	12
M92-C	C2	28*	T	51 (18)	62 (16)	11	0.22	—	13a
M93	C2	20*	T	33 (15)	48 (14)	15	0.22	37 (19)	13a
M94-C	C2	41*	IE(Low-g)	44 (11)	58 (16)	14	0.25	41 (15) ¤	13a
M-PD94-C	C2	21*	IE	44 (13)	63 (13)	19	0.34	—	13b
M-PD92a	C2	46*	IE	33 (15)	70 (12)	37	0.55	—	13b
M-PD92b	**C2**	**57***	**IE**	**30 (13)**	**73 (9)**	**43**	**0.61**	**55 (12)**	**13b**
M-PD93	C2	46*	IE	33 (14)	72 (10)	39	0.58	58 (8)	13b
M-PD94	C2	34*	IE	30 (10)	62 (13)	32	0.46	54 (15)£	13b
M-PD95a-C	C2	31*	IE	45 (14)	71 (13)	26	0.47	56 (15)¶	13b
M-PD95b-C	**C2**	**22***	**IE**	**50 (14)**	**82 (12)**	**32**	**0.64**	**64 (15)§**	**13b**
M-Co95c-C	C2	61*	IE	46 (7)	69 (7)	23	0.43	—	13b
PL92-C (2 sect.)	C4	24*	IE	48	76	28	0.54	—	14a
TO (8 cour.)	C2	61*	IE	35 (15)	62 (13)	27	0.42	—	15a
TO-C (5 cour.)	**C2**	**53***	**IE**	**43 (15)**	**77 (14)**	**34**	**0.60**	**70 (12)**	**15a**

Table Ic. Pre/post test data for 32 introductory *university* physics courses enrolling a total of N = 4832 students. Please refer to the heading of Table Ia for important explanations.

COURSE CODE [a] [1]	LEVEL [b] [2]	N [c] [3]	TYPE [d] [4]	PRE [e] [StdD] % [5]	POST [e] [StdD] % [6]	GAIN [f] G % [7]	NORM. GAIN [g] g [8]	MB [h] [StdD] % [9]	REF. [10]
ASU‡	U	82	T	37 (14)	53 (14)	16	0.25	—	1a
ASU1-C (4 cour.)‡	U	478	T	52 (15)	64 (15)	12	0.25	—	1a
ASU2-C	U	139	T	52 (19)	63 (18)	11	0.23	48 (15)	2a
ASU-HH-C‡	U	20*	IE	48 (17)	75 (13)	27	0.52	—	16
ASU-MW105-C	U	44	IE	36	68	32	0.50	43	2a
ASU-VH105-C	U	116	IE	34 (14)	63 (18)	29	0.44	61 (18)	2a,17a
CP-C	U	105	T	44 (19)	58 (21)	14	0.25	44 (14)	18a
CP-RK-Hon-C	**U**	**60**	**IE**	**59 (19)**	**84 (14)**	**25**	**0.61**	**69**	**18b**
CP-RK-Rega-C	U	70	IE	46	72	26	0.48	60	18b
CP-RK-Regb-C	U	69	IE	55	81	26	0.58	68	18b
EM90-C	U	121*	T	70??	78	8	0.27	67	19a,d
EM91-C	U	177*	IE	71	85	14	0.48	72	19a,d
EM93-C	U	158*	IE	70	86	16	0.53	73	19a,d
EM94-C	U	216*	IE	71	88	17	0.59	76	19a,d
EM95-C††††	**U**	**186***	**IE**	**67**	**88**	**21**	**0.64**	**76**	**19a,d**
IUpre93 (5 cour.)‡	U	346*	IE	44 (16)	74 (12)	30	0.54	—	20a,21,22a
IU93S†	U	154*	IE	37 (14)	73 (16)	36	0.57	55 (16)	22a
IU94S††	**U**	**166***	**IE**	**40 (17)**	**79 (14)**	**39**	**0.65**	**61 (16)**	**22a**
IU95S††	**U**	**209***	**IE**	**42 (15)**	**77 (15)**	**35**	**0.60**	**—**	**23**
IU95F†††	**U**	**388***	**IE**	**32**	**74 (18)**	**42**	**0.62**	**—**	**24**
Mich(De)-C	U	115*	IE	42	67	25	0.43	—	25
Mich(Ft)1-C	U	77	IE	47	67	20	0.38	—	26a
Mich(Ft)2-C	U	58	IE	45	65	20	0.36	—	26a
Mich (Ft)3	U	41	IE(Low-g)	39	53	14	0.23	—	26a
Mich (Ft)4	U	104	IE(Low-g)	31	47	16	0.23	—	26a
OS92-C	U	200#	T	48##	55##	7	0.13	—	17b
OS95-C	U	279*	IE	48	70 (20)	22	0.42	—	17b
UL94F-C	U	123*	T	44 (18)	54 (19)	10	0.18	—	27
UL-RM95S-C	U	119	IE(Low-g)	43 (18)	58 (21)	15	0.26	—	28a
UL-RM95Su-C	U	47	IE(Low-g)	44 (19)	58 (19)	14	0.25	—	28c
UML93-C	U	195*	IE(Low-g)	40	54	14	0.23	38	29
UML94-C	U	170*	IE(Low-g)	38	51	13	0.21	47	29

Table Ia,b,c Footnotes

a. CODE

AP: Advanced Placement
Hon: Honors
Reg: Regular (College Prep)
- C: Calculus-based
S: Spring Semester
F: Fall Semester
Su: Summer Session

Initials as "BC" of instructors are sometimes indicated. For full names and institutions see the references in column 10.

**According to ref. 2a (p. 147, top left) the two highest Arizona Honors FCI posttest scores (67% and 73%) are suspect and therefore those data are omitted from this tabulation.

‡ Mechanics Diagnostic test (36 questions) of ref. 1a was used (all others used FCI of ref. 2a or minor revisions - see below).

† Near original FCI: changes were made in three questions (12, 28, 29) to clarify the physics, but judging from subsequent analysis (see below) of similar slight changes, it is very doubtful that they affected the average pre- or posttest scores.

†† Very slightly revised FCI: changes were made in seven of the questions so as to remove possible ambiguities, but neither the scores nor the point biserial coefficients for those questions showed significant changes from the IUS93 test. Comparing the respective posttest results for IU93S (near-original FCI) and IU95S (very slightly revised FCI): $<S_i>$ = 37, 42; g = 0.57, 0.60; average point biserial coefficient = 0.38, 0.39; Kuder-Richardson reliability coefficient KR-20: 0.81, 0.81.

†††Slightly revised FCI (Form 072795 - 30 questions) almost identical to the 1995 revision (ref. 2b). Comparing the respective posttest results for IU93S (near original FCI) and IU95F (Form 072795): <Si> = 37, 32; g = 0.57, 0.62; average point biserial coefficient = 0.38, 0.44; Kuder-Richardson reliability coefficient KR-20: 0.81, 0.86.

†††† Revised 1995 FCI (30 questions, ref. 2b)

b. LEVEL

HS: High School
C2: 2-year College
C4: 4-year College
U: University

c. N: Number of students taking the posttest. An asterisk * means that the pretest average was determined from the pretest scores of only those students who took the posttest. When this is not done the error in the normalized gain g is probably less than 5% for courses with $50 \geq N \geq 20$, and probably less 2% for N > 50.

#: N was given as between 200 and 300 on the bar graph of the pre- and post-test scores (see "e" below - no other information was available).

d. TYPE

T: "Traditional" as defined in the Introduction.
IE: "Interactive Engagement" as defined in the Introduction.
Low-g means (<g>) < 0.3, as indicated in the Introduction.

e. PRE and POSTtest scores for the Force Concept Inventory (29 questions) of ref. 2a; except (where indicated by ‡ in column #1) the Mechanics Diagnostic test (36 question), or [where indicated by †'s (see footnote "a" above)] slightly revised versions of the FCI.

indicates that the pre- and post test scores were read from a FAX- transmitted bar graph (no other data were available) with an estimated total uncertainty of less than 5%, less than the usual standard deviations for such averages.

StdD: Standard Deviation. Both the test score and the StdD are given as a % of the total number of questions in the exam.

The "??" for the EM90-C pretest indicates that no pretest was given and the assumed score of 70% is based on pretest scores of similar later classes (EM91-C, 93-C).

f. %GAIN: $\%\langle \text{Gain} \rangle \equiv \%\langle \text{Posttest} \rangle - \%\langle \text{Pretest} \rangle \equiv \%\langle S_f \rangle - \%\langle S_i \rangle$, where <...> denotes an average over the entire class. Plots of <Gain> vs <Pretest> for high schools, colleges, and universities are shown in ref. 5a.

g. NORMalized GAIN <g>, defined as the ratio of the actual average gain to the maximum possible average gain, i.e., $\langle g \rangle \equiv \%\langle G \rangle / \%\langle G_{max} \rangle = (\%\langle S_f \rangle - \%\langle S_i \rangle) / (100 - \%\langle S_i \rangle)$, where <...> denotes an average over the entire class. For the graphical interpretation of <g> and its justification as a gauge of course effectiveness see ref. 5a.

h. %MB: average percentage score for the problem-solving Mechanics Baseline test (26 questions) of ref. 3. For a plot of MB score vs FCI posttest score and evidence that IE methods enhance problem-solving ability see ref. 5.

StdD: Standard Deviation. Both the test score and the StdD are given as a % of the total number of questions in the exam.

¤ ($N_{MB} = 46$), [Table Ib, $N_{MB} > N_{FCI}$ (column 3) because some students at Monroe Community College who took the MB did not take FCI pretest and therefore were not included in N_{FCI} (see footnote "c" above)].

£ ($N_{MB} = 38$), [Table Ib, see above.]

¶ ($N_{MB} = 37$), [Table Ib, see above.]

§ ($N_{MB} = 24$), [Table Ib, see above.]

B. Interactive-Engagement Methods

Table IIa,b,c shows the interactive engagement (IE) methods and materials that were most frequently employed by the 48 IE courses.

Table IIa. Interactive-engagement methods and materials used by the introductory *high-school* physics courses of Table Ia. The "•" indicates use, "?" indicates the presence of implementation problems as discussed in Sec. III. The **bold-faced** data indicate a normalized average gain (<g>) ≥ 0.6. Footnotes are placed after Table IIc.

COURSE	LEV-	N [c]	TYPE [d]	NORM.				METHODS				
CODE [a]	EL [b]			GAIN [e]	Coll.Peer	MBL	Concept	OCS	Mod-	SDI	Other	PER [m] -based
				g	Inst. [f]	[g]	Tests [h]	ALPS [i]	eling [j]	[k]	Methods [l]	Text or No Text
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
AZ-AP	HS	33	T	0.27	•?	•?	—	—	•?	—	—	—
AZ-Hon (2 cour.)**	HS	62	T	0.25	•?	•?	—	—	•?	—	—	—
AZ-Reg (18 cour.)	HS	612	T	0.29	•?	•?	—	—	•?	—	—	—
BC-Hon	**HS**	**22***	**IE**	**0.69**	•	•	—	—	—	—	**n,o,p,q**	—
BC-Reg	HS	43*	IE	0.34	•	•	—	—	—	—	o,p,q	—
Chicago-Reg	HS	56	T	0.21	—	—	—	—	—	—	—	—
CL-Reg	HS	20*	IE	0.42	•	•	—	•	—	•	n,o,p,q,r,s,t	—
ELM-Hon	**HS**	**20***	**IE**	**0.68**	•	•	—	—	•	—	**u**	**Dekker (ref. 61)**
GS-Hon	HS	63	IE	0.53	•	•	—	—	•	—	—	—
GS-Hon95	**HS**	**49**	**IE**	**0.61**	•	•	—	—	•	—	—	—
LT-Hon	HS	27*	IE	0.57	•	•	—	•	•	•	p	None
MW-Hon	**HS**	**30**	**IE**	**0.62**	•	•	—	—	•	—	—	—
RM94Reg	HS	38*	IE	0.48	•	•	—	—	—	—	o,p,q,v,w,x,y	—
RM95Reg	HS	38*	IE	0.51	•	•	—	—	—	—	o,p,q,v,w,x,y	—

Table IIb. *Interactive-engagement* methods and materials used by the introductory *college* physics courses of Table Ib. Please refer to the heading of Table IIa for important explanations.

								METHODS				
COURSE CODE [a]	LEVEL [b]	N [c]	TYPE [d]	NORM. GAIN [e] g	Coll.Peer Inst. [f]	MBL [g]	Concept Tests [h]	OCS ALPS [i]	Modeling [j]	SDI [k]	Other Methods [l]	PER [m] -based Text or No Text
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
DB-C	C4	27*	IE	0.36	•	•	•	•	—	—	n	—
LF92	C2	25	T	0.20	—	—	—	—	—	—	—	—
M92-C	C2	28*	T	0.22	—	—	—	—	—	—	—	—
M93	C2	20*	T	0.22	—	—	—	—	—	—	—	—
M94-C	C2	41	IE(Low-g)	0.25	•?	•?	—	—	—	—	—	—
M-PD94-C	C2	21*	IE	0.34	•	—	•	•	•	—	n, t, w, x, z	D’Ale. (ref.13e)
M-PD92a	C2	46*	IE	0.55	•	—	•	•	•	—	n, t, w, x, z	D’Ale. (ref.13e)
M-PD92b	**C2**	**57***	**IE**	**0.61**	•	—	•	•	•	—	**n, t, w, x, z**	**D’Ale. (ref.13e)**
M-PD93	C2	46*	IE	0.58	•	—	•	•	•	—	n, t, w, x, z	D’Ale. (ref.13e)
M-PD94	C2	34*	IE	0.46	•	•	•	•	•	—	n, t, w, x, z	D’Ale. (ref.13e)
M-PD95a-C	C2	31*	IE	0.47	•	•	•	•	•	—	n, t, w, x, z	D’Ale. (ref.13e)
M-PD95b-C	**C2**	**22***	**IE**	**0.64**	•	•	•	•	•	—	**n, t, w, x, z**	**D’Ale. (ref.13e)**
M-Co95c-C	C2	61*	IE	0.43	•	•	•	•	•	—	n, t, w, x, z	D’Ale. (ref.13e)
PL92-C (2 sect.)	C4	24*	IE	0.54	•	•	—	—	—	—	—	—
TO (8 cours.)	C2	61*	IE	0.42	•	•	—	—	—	—	t	—
TO -C (5 cours.)	**C2**	**53***	**IE**	**0.60**	•	•	—	—	—	—	**t**	—

Table IIc. *Interactive-engagement* methods and materials used by the introductory *university* physics courses of Table Ib. Please refer to the heading of Table IIa for important explanations.

COURSE	LEV-	N [c]	TYPE [d]	NORM.				METHODS				
CODE [a]	EL [b]			GAIN [e] g	Coll.Peer Inst. [f]	MBL [g]	Concept Tests [h]	OCS ALPS [i]	Mod-eling [j]	SDI [k]	Other Methods [l]	PER [m] -based Text or No Text
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
ASU [j]	U	82	T	0.25	—	—	—	—	—	—	—	—
ASU1-C (4cour.)‡	U	478	T	0.25	—	—	—	—	—	—	—	—
ASU2-C	U	139	T	0.23	—	—	—	—	—	—	—	—
ASU-HH-C‡	U	20*	IE	0.52	•	—	—	—	•	—	—	—
ASU-MW105-C	U	44	IE	0.50	•	•	—	—	•	—	—	—
ASU-VH105C	U	116	IE	0.44	•	—	—	•	—	•	—	—
CP-C	U	105	T	0.25	—	—	—	—	—	—	—	—
CP-RK-Hon-C	**U**	**60**	**IE**	**0.61**	•	•	**—**	•	•	**—**	**—**	**Knight (ref. 18c)**
CP-RK-Rega-C	U	70	IE	0.48	•	•	—	•	•	—	—	—
CP-RK-Regb-C	U	69*	IE	0.58	•	•	—	•	•	—	—	Knight (ref. 18c)
EM90-C	U	121*	T	0.27	—	—	—	—	—	—	—	—
EM91-C	U	177*	IE	0.48	•	—	•	—	—	—	—	—
EM93-C	U	158*	IE	0.53	•	—	•	—	—	—	a'	—
EM94-C	U	216*	IE	0.59	•	—	•	—	—	—	a'	—
EM95-C††††	**U**	**186***	**IE**	**0.64**	•	**—**	•	**—**	**—**	**—**	**a'**	**—**
IUpre93 (5 cour.) ‡	U	346*	IE	0.54	•	—	—	—	—	•	q,b',c'	—
IU93S†	U	154*	IE	0.57	•	•	•	—	—	•	q,b',c',f'	—
IU94S††	**U**	**166***	**IE**	**0.65**	•	•	•	**—**	**—**	•	**q,b',c',f'**	**Reif (ref. 62)**
IU95S††	**U**	**209***	**IE**	**0.60**	•	•	•	**—**	**—**	•	**b',c',d'**	**—**
IU95F†††	**U**	**388***	**IE**	**0.62**	•	•	•	**—**	**—**	•	**x,b',d',h'**	**—**
Mich(De)C	U	115*	IE	0.43	•	•	•	•	•	—	g'	—
Mich(Ft)-1C	U	77	IE	0.38	•	•	—	—	—	—	—	—
Mich(Ft)-2C	U	58	IE	0.36	•	•	—	—	—	—	—	—
Mich (Ft)3	U	41	IE(Low-g)	0.23	•?	•?	—	—	—	—	—	—
Mich (Ft)4	U	104	IE(Low-g)	0.23	•?	•?	—	—	—	—	—	—
OS92-C	U	200	T	0.13	—	—	—	—	—	—	—	—
OS95-C	U	279*	IE	0.42	•	•	—	•	—	•	t,x,z,e',g',i'	—
UL94F-C	U	123*	T	0.18	—	—	—	—	—	—	—	—
UL-RM95S-C	U	119	IE(Low-g)	0.26	•?	•?	—	—	—	—	—	—
UL-RM95Su-C	U	47	IE(Low-g)	0.25	•?	•?	—	—	—	—	x,h'	—
UML93-C	U	195*	IE(Low-g)	0.23	•?	—	•?	—	—	—	—	—
UML94-C	U	170*	IE(Low-g)	0.21	•?	—	•?	—	—	—	—	—

Table IIa,b,c Footnotes

a - d. Same as Table I.

e. NORMalized GAIN <g>, defined as the ratio of the actual average gain to the maximum possible average gain, i.e., $<g> \equiv \%<G>/\%<G_{max}> = (\%<S_f> – \%<S_i>) / (100 - \%<S_i>)$, where <...> denotes an average over the entire class.

f. Collaborative Peer Instruction (CPI): see, e.g., ref. 30 - 32 and references therein. CPI is an integral part of many IE methods, e.g., MBL [as employed in TST (ref. 33b), RTP (ref. 33c,e), Workshop Physics (ref. 14), and "Targeted MBL Tutorials" TMT (ref. 34)]; Concept Tests; OCS/ALPS; Modeling; SDI; and "McDermott Recitation Tutorials" (MRT - ref. 35).

g. Microcomputer-Based Laboratories (MBL): Ref. 33. MBL is an integral part of Workshop Physics (ref. 14c-e); Tools for Scientific Thinking (TST) (ref. 33b); Real Time Physics (RTP) (ref. 33c,e); "Targeted MBL Tutorials" (TMT) (ref. 34); and is utilized in several SDI labs. For commercially available MBL equipment see ref. 33d. Most of the MBL use in the courses of Table II is done within the context of one of the above methods or similar strategies devised by the instructors.

h. Concept Tests: Ref. 19b-e, 36.

i. Overview Case Studies (OCS) and Active Learning Problem Sets (ALPS): Ref. 17c,d,f.

j. Modeling Instruction: Refs. 16, 37, 38.

k. Socratic Dialogue Inducing Labs (SDI): Refs. 20-22, 39-42.

l. Only methods and materials for mechanics instruction are listed.

m. PER ≡ Physics Education Research (see, e.g., refs. 43-46); for a review of PER-based texts see ref. 43i.

n. Hand-held graphing calculators. Advanced models allow the graphing of data, see e.g., ref. 47. Such systems as Calculator Based Laboratories (CBL) from Texas Instruments allow many MBL-type activities, see <http://www.ti.com/calc/>.

o. Physics Academic Software, see, e.g., ref. 48.

p. Video disks, see, e.g., ref. 49.

q. Mechanical Universe Video, see, e.g., ref. 50.

r. InfoMall, see, e.g., ref. 51 and <http://bluegiant.phys.ksu.edu>.

s. PRISMS (Physics Resources and Instructional Strategies for Motivating Students), see, e.g., ref. 52.

t. Ranking-task questions, ref. 53.

u. Interactive video, see, e.g., ref. 54.

v. ALPS and OCS (see "i" above), Modeling (see "j" above), and SDI (see "k" above) influence design of activities but are not specifically used by the students.

w. modest use of some parts of *Physics by Inquiry,* L. C. McDermott *et al*., ref. 55.

x. *Interactive Physics,* a product of Knowledge Revolution, <http:www.krev.com/>, see also ref. 56.

y. Construction contests (e.g., egg-drop, catapult).

z. Goal-less Problems, ref. 13c,e

a'. Classtalk (ref. 57) provides hand-held computers for students, a master computer for the instructor, and a classroom network which allows immediate feedback from Concept Tests or a lecturer’s questions.

b'. Minute Papers, ref. 41e, 58.

c'. DIagnostic Student COmputerized Evaluation (DISCOE), ref. 21, 59.

d'. *First Class* (ref. 60) allows electronic-bulletin-board discussions, file sharing, and collaboration among students and instructors. It is available from SoftArc Inc., 100 Allstate Parkway, Markham, Ontario, Canada, L3R6H3. See <http://www.softarc.com>.

e'. Context-Rich Problems (ref. 31and <http://www.physics.umn.edu/groups/physed/index.html>).

f'. Out-of-Lab Problems (ref. 22d).

g'. Experiment Problems (ref. 17e,f).

h'. MBL lecture demonstrations in "lecture."

i'. Interactive simulations in "lecture."

Tables IIa,b,c show that the ranking of the more popular IE methods in terms of number of IE courses using each method is: Collaborative Peer Instruction (CPI),[63] 48 (*all* courses); Microcomputer-Based Labs (MBL), 35; Concept Tests, 20; Modeling, 19; Active Learning Problem Sets (ALPS) or Overview Case Studies (OCS), 17; physics-education-research based text or no text, 13; and Socratic Dialogue Inducing (SDI) Labs, 9. [For simplicity, courses combined into one "course" [TO (8 courses), TO-C (5 courses) and IUpre93 (5 courses) are counted as one course each.] The ranking in terms of number of students using each method is: Collaborative Peer Instruction (CPI), 4458 (*all* students); MBL, 2704; Concept Tests, 2479; SDI, 1705; OCS/ALPS, 1101; Modeling, 885; research-based text or no text, 660.

It should be emphasized that the above rankings are by *popularity* within the present survey, and have no necessary connection with the *effectiveness* of the methods relative to one another. In fact, it is quite possible that some of the "Other Methods" referenced in column 12 of Table II could be more effective than any of the more popular strategies. The tabulations and references in Table II enable teachers and researchers ready access to the literature and materials relevant to all the IE methods of the survey.

Because details of IE-method implementation are important (Sec. III) and appear to account for much of the spread in the <g> values of IE courses,[5a] it is worthwhile to consider not just the methods themselves, but also other factors. Tables I and II show that IE methods are being used (a) in many different types of institutions (selective and non selective, small- and big-city high schools, two-year and four-year colleges, research universities), (b) for diverse student groups (in-need-of-remediation, regular, honors, science, non-science, engineering); and (c) by instructors who, for the most part, are active contributors to the physics-education literature.

Since the present survey suggests that use of the IE methods of Table II can increase mechanics course effectiveness well beyond that achieved with T methods, *the methods would appear to deserve serious consideration by physics teachers who wish to improve their courses, by physics-education researchers who may wish to test or formulate general ideas on instruction or learning, by creators of instructional materials, and by designers of new introductory physics courses.*[64]

Several features of the IE methods are noteworthy:

1. Interdependence, Mutual Compatibility, Electronic Availability

IE methods are usually interdependent (see Table II, footnotes f and g). As demonstrated in Table II and refs. 40 & 42b, they are mutually compatible and can be melded together to enhance one another’s strengths and modified to suit local conditions and preferences (especially easy if materials are available electronically[13e,19d,22d] so as to facilitate copying, pasting, and cutting). In addition to allowing easier modification and mixing of materials by instructors, electronic availability has the added virtue of allowing continual and needed improvement of IE methods and materials, in accord with a *redesign process* (described by Wilson and Daviss[65]) of continuous *long-term* classroom use, feedback, assessment, research analysis, and revision. Adjustments and updating of educational materials in accord with electronic feedback from users can be made within days rather than years. Of course, great care must be taken not to compromise teacher's guides, answer sheets, and standardized tests by making them generally available at non-protected sites.

2. Emphasis on Problem Solving

Most of the IE courses of Tables I and II emphasize problem-solving in addition to conceptual understanding. In most IE courses some of the problem solving requires critical thinking and mathematical skill as well as the understanding of concepts.[66] Thus it may not be surprising that an analysis[5a] of the problem-solving Mechanics Baseline data of Table I suggests that IE methods *enhance* problem-solving ability.

C. "Effectiveness" Defined

In the introduction it was stated that the present survey strongly suggests that classroom use of IE methods can increase mechanics course "effectiveness" in both conceptual understanding and problem-solving *well beyond that achieved with T methods*.

But are the IE methods of this study "effective" in some absolute sense? First it should be emphasized that (a) "the FCI was developed to assess the effectiveness of mechanics courses in meeting a *minimal performance standard* : to teach students to reliably discriminate between the applicability of scientific concepts and naive alternatives in common physical situations" [37c] (our *italics*); (b) the Mechanics Baseline test is "the next step above the inventory in mechanics understanding ...(and).... emphasizes concepts that cannot be grasped without formal knowledge about mechanics."[3] Thus these tests do not pretend to measure advanced mechanics competence, but rather only a minimal facility which might be hoped for at the end of an introductory course.

Among desirable outcomes of the introductory course that the tests *do not measure directly* are e.g., students' (a) satisfaction with and interest in physics; (b) understanding of the nature, methods, and limitations of science; (c) understanding of the processes of scientific inquiry such as experimental design, control of variables, dimensional analysis, order-of-magnitude estimation, thought experiments, hypothetical reasoning, graphing, and error analysis; (d) ability to articulate their knowledge and learning processes; (e) ability to collaborate and work in groups; (f) communication skills; (g) ability to solve real-world problems; (h) understanding of the history of science and the relationship of science to society and other disciplines; (i) understanding of, or at least appreciation for, "modern" physics; (j) ability to participate in authentic research. It can be argued that some outcomes "a" - "g" (e.g., "b"[73a]) are more likely to have been achieved by students who do well on the FCI/MD and MB tests. Nevertheless, because evidence for these outcomes cannot be directly offered by such testing, and because most instructors would regard at least some of "a" - "j" to be important objectives of the introductory course, the FCI/MD and Mechanics Baseline test scores should not, in my opinion, be uncritically taken to measure the general effectiveness or success of a course. They can, however, be taken *to measure effectiveness in the narrow sense of the attainment of minimal competence in mechanics.* Most instructors would probably agree that this should be a prime objective of an introductory mechanics course. The 48 interactive-engagement courses of this study appear, on average, to be much more effective in this minimal sense than traditional courses. But even in this minimal sense, none of the courses is in the High-g region and some are even in the Low-g region characteristic of traditional courses. Thus, in absolute terms, *the IE methods of this study could all stand improvement and more work seems to be required on both their content and implementation.*

Are some IE methods more effective than others? Within the context of this study, such comparison is somewhat uncertain because (a) most of the IE courses of the survey employed various IE methods in combination (Table II) making intercomparison of individual methods difficult, and (b) there are uncontrolled variables such as the characteristics of the instructors, students, institutions, and implementations which lead to large spreads in <g> for courses using similar methods. Although future more refined studies may be able to rank the effectiveness of different methods with respect to one another and with respect to particular course objectives, the crucial outcome of the present survey is that *all* the methods which are more effective than the traditional in promoting minimal competence in mechanics were "*designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors*." At this stage (a) the particular method used by an instructor may be less important than the skill of that instructor in promoting effective interactive engagement of students, (b) teachers might be well advised to try first those methods which best match their own inclinations, course objectives, teaching styles, students' characteristics, and resources.

D. Are "Good" Teachers Sufficient For High Quality Physics Instruction?

In his Millikan award paper, David Griffiths[67] wrote: "But I believe .....(physics).... enrollments would have held up in spite of all these influences....(poor job market, unsupportive environment for students, lack of science-course distribution requirements)....were it not for the abysmal quality of physics instruction, especially at our large research universities....In my opinion by far the most effective thing we can do to improve the quality of physics instruction - much more important than modifications in teaching technique - is to hire, honor, and promote good teachers." Although few would deny that good teachers are *necessary* for high quality instruction, that they are not *sufficient* is suggested by the persistent placement of T courses in the Low-g range, even when conducted by highly regarded teachers.[1a,19]

A referee, taking a stance similar to that of Griffiths, implied that some or all of the IE instructors who achieved <g>'s above the traditional range might also have achieved similar high <g>'s using T methods. Here are four counter examples, all drawn from Table I:

1. *David Hestenes*

Hestenes was one of the four professors in course ASU1-C of Table Ic. That course used traditional methods to attain <g> = 0.25. Here <g> represents a number-of-student weighted average for four courses which all achieved similar <g>'s in the low-g region. Then in course ASU-HH-C of Table Ic, Hestenes teamed with Halloun and employed the Modeling method to obtain <g> = 0.52.

2. Eric Mazur

In course EM90C, Mazur employed traditional methods in a course that attained <g> = 0.27. After he switched to Concept Tests[19b] his successive courses EM91C, EM93C, EM94C, and EM95C achieved, respectively, <g> = 0.48, 0.53, 0.59, and 0.64.

3. Malcolm Wells

As recounted in ref. 38, the late Malcolm Wells, inspired by PSSC and Harvard Project Physics workshops in the 70's, had abandoned the traditional lecture-demonstration method early in his career. In the 80's he regularly employed an inquiry approach based on the Karplus learning cycle.[68] However, when he administered the MD he discovered that his course's performance was about the same as characterizes current T courses. When Wells shifted to the Modeling method his courses MW-Hon of Table Ia, and ASU-MW105-C of Table Ic achieved, respectively, <g> = 0.61 and 0.50.

4. Richard Hake

In 1980 (prior to publication of the MD[1a]), I used T methods in a course for prospective elementary teachers.[20b] Because these students were extremely weak in mathematics, my first exam consisted of conceptually oriented multiple-choice questions quite similar to those of the FCI/MD. The test scores were abysmal and showed that my brilliant lectures and exciting demonstrations had passed through the students' minds leaving no measurable trace. On the advice of Arons, I started using a Socratic approach in the labs and test results improved considerably. Later on, in pre-med-type courses, I used Socratic Dialogue Inducing Labs plus more interactive lectures and recitations in courses IUpre93 and IU93S of Table Ic, which achieved <g> = 0.54, 0.57.

III. IMPLEMENTATION PROBLEMS

Reif[45a] has discussed the nature and seriousness of implementation difficulties, drawing a parallel between hypothetical situations in health care and physics education. Suppose medical science has produced pills which can reliably cure all diseases, but people do not take the pills because they believe in folk medicine or natural healing, or else simply refuse to follow the recommended pill-taking regimen. Then practical medical implementation will have failed. Likewise:

> "....practical educational implementations face similar difficulties (quite apart from motivational factors). For example, suppose that we understood perfectly the thought and learning processes needed for physics. All this understanding would still be insufficient to ensure practical educational implementation if students have misleading beliefs about science or do not actually engage in the recommended learning activities....
>
> An introductory physics course needs....to discuss explicitly the goals of science and the ways of thinking useful in science....They need to be constantly kept in students' focus, and be used as a framework within which more specific scientific knowledge and methods are embedded....
>
> Even the best instructional materials and methods are useless if students do not actually engage in the recommended learning activities. This is well recognized in efforts designed to train good athletes or musical performers. There coaches or teachers provide very frequent supervision, with the guidance and feedback necessary to ensure that students acquire good habits – and to prevent bad habits which may be difficult to break or even lead to injuries.... *How then can one provide students with adequate guidance and feedback in practical contexts dealing with many students? In my judgment, this is a fundamentally important problem which, if left unsolved, will remain a bottleneck hindering even the best designed instruction.*" (Our *italics*.)

As many physics instructors have discovered, it is one thing to learn about apparently successful pedagogical methods from talks, articles, or workshops, but quite another to implement them successfully in the classroom. The problems indicated above by Reif are (a) lack of student motivation[71,72] (especially severe for students in IE courses who dislike any departure from the traditional methods to which they have become accustomed and under which their grades, if not their understanding, may have flourished[9,14c,19b,26,32,71]); (b) naive student beliefs about the nature of science and learning[72,73]; and (c) the difficulty of providing adequate coaching[74,75] and practice[75] for students (and, I might add, instructors) in large-enrollment classes. In addition, there are (d) difficulties in integrating multi-component courses[76]; (e) poor science[77] and math preparation[78] of students (reflecting in part the dismal failure of colleges and universities to adequately assist in the preparation of precollege teachers[79-81]); and (f) organizational problems[82] such as: inertia, bureaucracy, inadequate funding, lack of enthusiasm for non-physics-major education, grade inflation, the administrative misuse of student evaluations to gauge the cognitive (rather than just the affective) impact of courses, and the indifference or animosity of colleagues and administrators towards new instructional methods.

The use of interactive-engagement methods appears to be necessary but not sufficient for marked improvement over traditional methods. Seven of the IE courses of Tables I and II are in the Low-g range of the traditional courses. In order to benefit from past experience it seems worthwhile to discuss these IE(Low-g) cases in some detail in the constructive spirit of the *redesign process,*[65] especially because personal experience with the Indiana courses and communications with most of the IE instructors in this study suggest that similar though less severe implementation problems were common.

> *One does not get anywhere simply by going over the successes again and again, whereas by talking over the difficulties people can hope to make some progress.*
>
> Paul Adrien Maurice Dirac

A. Case Studies

1. Arizona High Schools [AZ-AP, AZ-Hon, and AZ-Reg of Tables Ia and IIa]

In Tables I and II, the high-school courses[2a] AZ-AP, AZ-Hon, and AZ-Reg are listed as traditional because pre/post test data for the pre-workshop courses (traditional) and for the post-workshop courses (IE methods attempted) *were not significantly different* and were averaged together to yield the one set of pre/post data reported in ref. 2a and quoted in Table 1. The *post*-workshop courses, by themselves, serve as examples of "IE(Low-g)" courses. These courses were evidently beset with an implementation problem: "From discussions with the teachers after the....(post-workshop courses).... it has become clear that they were so involved with the mechanics of the method - computers, lab activities, discussion technique - that they failed to fully appreciate the crucial pedagogical core....(modeling as the method of science[16,37,38]).... that makes it effective."[2a] My experience[20a,22c] suggests that in addition to partaking in workshops, instructors new to IE methods need to serve *apprenticeships* under experienced and effective IE teachers.

In Table IIa, the methods employed in the AZ-AP, AZ-Hon, and AZ-Reg courses are indicated by "•?" to indicate the presence of an implementation problem.

2. University of Massachusetts at Lowell [UML93-C, UML94-C of Tables Ic and IIc]

UML93-C and UML94-C represent an attempt to carry over the Concept Test method used successfully by Mazur at Harvard (average FCI pretest scores $<S_i> \approx 70\%$), to less well-prepared students at the University of Massachusetts at Lowell ($<S_i> \approx 39\%$). Approximate versions of the Mazur method have been successfully transported to low $<S_i>$ classes as shown in Table II for thirteen courses: DB-C, M-PD94-C to M-Co95c-C, and IU93S to IU95F.[36] However, at UML, although the students greeted Concept Tests with great enthusiasm and interest,[29] the implementations lacked certain crucial Harvard features. At Harvard[19a,c] students *take a quiz at the start of each lecture* on the reading assignment. Grades on these quizzes reduce the final exam weight. Failure to participate in the Concept Tests voids such quiz points. Thus at Harvard there is a direct grade incentive for coming into the lecture prepared to consider the physics of the Concept Tests and there is an indirect grade incentive to participate in them. However, at UML no direct or indirect grade incentives of the Harvard-type are given, and "the Concept Tests end up taking time away from the lecture and this time is not made up by students on their own time (as it is at Harvard)." [19a]

For Concept Tests to be successful at Indiana University, it has been necessary to provide a direct grade incentive: the group scores count between 12 and 15% of the final course grade. Thus students are motivated to come to class prepared to consider the physics of the Concept

Tests and to take them seriously. Had grade incentives been offered at UML, the pre-post test results might have been more encouraging. Of course, it is possible that the UML results represent an improvement over conventional introductory courses at UML, as occurred at the University of Louisville (Case 4 below), but unfortunately there is no UML baseline data.

3. University of Michigan at Flint **[**Mich(Ft)3,4 of Tables Ic and IIc]

Don Boys wrote: "The person giving the lectures also supervised the labs but did not always have the agreement of the lab instructor as to the worth of that style....(Tools for Scientific Thinking)....of teaching. In the Fall of 1993 almost all ....Mich(Ft)4.... labs were taught by someone who was writing his thesis and totally unfamiliar with the style of labs. We encouraged him to visit our labs to see how we did things but that did not occur."[26]

Here it would appear that the lab instructor failed to serve as a coach[74,75] and to provide students with adequate guidance and feedback. My experience[20a,22c] has been that it is essential to educate and carefully supervise lab instructors who are new to IE methods. At Indiana we have had reasonable success using an apprenticeship method (cf. the assistant coach in athletics) in which new instructors serve as assistants to experienced and successful instructors for at least a semester.

According to Don Boys "the students .... in the Mich(Ft)3,4 courses....are taking the course because it is required for some health related profession. They are poorly prepared, afraid of math ....and regard physics as the enemy."

Poor preparation of incoming students seems to afflict most physics introductory courses including those at Indiana University.[77,78] Aside from raising admission standards, there seems to be little that can be done about this in the short term, although incoming diagnostic tests[83,84c] may be helpful in early recognition and positive intervention for potential low-gain students. As for fear of physics and anti-physics attitudes, the nature of science and learning needs to be explained and emphasized throughout the course.[19b,43d,45a,62,72,73]

4. University of Louisville **[**UL-RM95S-C, Spring 1995) of Tables Ic and IIc]

With regard to UL-94F-C given in the Fall of 1994, Roger Mills wrote " I gave the Hestenes FCI last semester in the old lab format. The scores were....(see Table Ic, UL94F-C, g = 0.18).... I hope the new labs will improve on that, but better use of the lecture-recitation is likely to be important too."[27]

"You commented that the FCI results were unlikely to be seriously affected by the labs alone. We are now using a variant of the Real Time Physics approach in the lab. The....scores we have from this semester are....(see Table Ic, UL-RM95S-C, g = 0.26)....I'll give the test results to the lecture instructors, but I doubt that the results will be given much attention......Most of the people here are convinced that they are doing a great job of teaching, and if the FCI indicates otherwise, then there must be something wrong with the FCI. They do not hear well when told of its wide use and testing, and they are largely unaware of the considerable attention.... (that's been given)....to improve student participation in the learning process. A concentration on pedagogy is thought to be a lesser activity in comparison with the important research which they are conducting. In distributing rewards, good student reviews carry far more weight than do innovations in the classroom."[28a]

Although at Louisville, <g> increased by 44%, possibly due to the Real Time Physics (RTP)[33c,e] labs, <g> (UL-RM95S-C) is still close to the 0.23 average of traditional courses.

According to Mills,[28a,c] it would appear that the execution of the RTP labs at UL was substantially in accord with the recommended guidelines. In some respects, Mills went beyond the guidelines to try to make the RTP labs more effective:

"Each TA has done the experiments under my direct supervision, accompanied by questions intended to emphasize important aspects and to cue them about points which they should attend to with their own students. At the beginning of each semester, a separate period is set aside to familiarize the students with the equipment. This is very successful in bringing the students to a point where their use of the equipment is easy and comfortable, regardless of gender or minority group circumstances which may have resulted in lack of prior experience. The students usually work in groups of two or three. ....

We have found the use of RTP to be positive, and some of the students have gone to our department chair to complain that the next (E&M) lab is not as progressive."

According to Mills,[28a] there were, however, two possibly significant departures from the RTP-recommended guidelines at UL:

(1) The exercises intended as lab-followup homework assignments were used at UL as prelab exercises. This prelab preparation counted towards the lab grade and brought students into the lab with objectives and pertinent physics more clearly in mind, thus allowing "timely completion of most of the lab exercises" in two hours rather than the three hours informally recommended[28b] by one of the designers of RTP. But according to Laws,[14b] good RTP instructors (a) "require that followup homework assignments...(that review the observations in the labs)... be completed in the labs. Each assignment takes students about 20-30 minutes to complete and is collected at the beginning of the next lab period,....(and)....(b) "lead a *discussion of the homework* at the beginning of the next lab period with the students. In doing similar lab work in the Workshop Physics course, I have found that these steps are absolutely essential. UL's change from the RTP-recommended homework procedure could have been detrimental to student learning."

(2) Due to space constraints, the valuable RTP Lab #10 (students apply taps to balls rolling on a horizontal floor so as to simulate projectile motion) had to be omitted.

5. University of Louisville **[**UL-RM95Su-C, Summer 1995) of Tables Ic and IIc]

Roger Mills wrote: "In the following summer session..... I taught both the lecture and the lab. We used the Real Time Physics labs, and I used some MBL materials as part of the lecture. There were about 55 people in the lecture, so we didn't try to use discussion clusters in lecture. I also used *Interactive Physics* II as part of the demonstrations. (For what it's worth, the class grades were the highest that I have ever had for a class that size, and even the anonymous student evaluations were warm and glowing for a change.) I did use the FCI in the labs. With 47 people being tested....(See Table Ic, UL-RM95Su-C, g = 0.25)."

"I have thought further about the near-consistency of our Spring 1995 and Summer 1995 results. Although in Summer 1995 I used demonstration aids which were not included in the Spring 1995 courses, I did not directly engage the students as you might have with SDI. That the improvement in conceptual development was either no better or even a little worse may reflect the fact that implementation was equally impaired in both courses. The gain was nearly the same. This would underscore your point about the crucial importance of the active engagement in the lecture-recitation part of the course. About 80% of my tests related to problem solving, and only 20% related to conceptual development, through T-F questions. Thus I did not really motivate my students by that means to improvements in conceptual understanding.

You are correct that the lecture exams made no attempt to test for anything covered in the labs in either Fall ’94, Spring ’95, or even in Summer ’95. Since the enrollment in the lecture course does not actually require enrollment in the labs, not all of the students in the lectures were enrolled in the labs. There was a difference of 11 students, and I couldn’t expect those persons to be responsible for instruction which they hadn’t encountered."

6. Monroe Community College[13] **[**M94-C of Tables Ib and IIb]

Paul D’Alessandris writes: "That MBL is not a *sufficient* condition for achieving high g is shown by the fact the grafting of MBL onto the traditional lecture course M94-C resulted in g = 0.25. I taught the lab for many of these students and we used RTP by the book. In the Fall of 1995 a course similar to M94-C was repeated, that is, it was a traditional calculus-based course except for the fact that an RTP lab was substituted for the standard lab. Identical results were obtained: g was again 0.25. In that section only 8 students completed the class....(it is not listed in Tables I, II because N < 20)."

" ’Using RTP by the book’ means using the RTP materials after attending a day-long workshop in Orlando and two 3-day workshops organized by O’Kuma and Hieggelke, ...(holding the 3-hour labs)....in a laboratory space completely redesigned for MBL, collecting, grading and discussing all homework, collecting and commenting on all activity packets,and using both the FCI and FCME as diagnostics instruments. Since 1994, I have taught one or two lab sections and two or three other professors have taught the remaining sections. Although the other professors have not attended workshops, they followed the protocol outlined above. The only negative aspect to the implementation, aside from the other instructors lack of officially sanctioned workshop attendance, has been the lack of assistants to help run the lab. The labs are run with 24 students and one instructor. However, two-year college instructors as a rule are used to being overworked in the laboratory. I do not believe that the results I have reported are the result of an implementation problem.....(in so far as the conduct of the lab itself is concerned)...."

"While searching through FMCE ....(Force Motion Concept Evaluation of Thornton and Sokoloff[33e])...... records, I have yet to find a section whose postest FMCE average was below 65%, which is comparable to the FMCE posttest average informally reported[14b] for 18 Dickinson-Workshop-Physics students. More commonly, the FMCE average is in the 70’s, 80’s, and occasionally 90’s. In a nutshell, MBL has been well implemented ....(within the labs themselves).... at MCC, as evidenced by relatively high FMCE scores. In addition, student and other faculty response has been very positive, even in the face of some initial trepidation."

"My personal belief is that students learn the interrelationships between kinematic variables much better through MBL than through most paper/pencil activities. They also are orders of magnitude more fluent with graphs. However, for the two semesters in which some students had MBL grafted onto a traditional course, their FCI gains were unaffected. I had many of these students in lab; I believe many were killing time. They enjoyed doing the experiments, but the experiments didn’t connect with the course as a whole. Somehow this prevented these students from fully assimilating the concepts basic to the lab work."

"The prominent influence of the non-lab part of a course can be shown in another way. In the 1995-Spring-semester courses MPD95a-C (g = 0.47) and MPD95b-C (g = 0.64), I conducted all the interactive lectures using my Spiral workbook and both courses utilized RTP labs.... (The g-difference in these two courses is evidently due to the fact that the former is a day course while the latter is a night course drawing older and more motivated students- see PD's quotes in ref. 5a).... The two subsequent calculus-based courses given in the 1995 Fall semester again both utilized interactive SPIRAL-workbook lectures and RTP labs (all taught by instructors who had taught RTP labs the previous year). However, there were marked differences in the lecture parts of those courses. In one course .... [MPD95c-C (not shown in Tables I and II because N = 15)]...... I gave interactive lectures and achieved g = 0.63. In the other course .... (MCo95d-C of Tables I and II)].... my colleagues gave the interactive lectures and achieved g= 0.43. My colleagues were new to the Spiral-workbook lecture method. The two courses had nearly *identical* MBL labs, but had lectures taught by instructors with different levels of proficiency. (I do not mean to demean the job that my colleagues did. In fact, their first year using the Spiral-workbook was better than my first year using it with engineers (M-PD94-C, g = 0.34)."

"That MBL is not a *necessary* condition for achieving high-g is shown by the fact that high g's were obtained at Monroe *without* MBL: M-PD92a (g = 0.55), M-PD92b (g = 0.61), M-PD93 (g = 0.58). Although my gut feeling is that MBL with traditional lectures helps, my data suggest that its effect is minimal. I think MBL in conjunction with an attempt at IE lectures (my first teaching engineers as well as my colleagues first try at IE) can get g ~ 0.40. MBL plus 'well-executed' IE can get you 0.50 or more. Of course, when I taught the non-calculus courses (M-PD92a, M-PD92b, M-PD93) without MBL, I also got 0.50 or more, although I had the opportunity to spend more time on building conceptual understanding in non-calculus physics than with the engineers in calculus physics. I would like to see what I could do in non-calculus physics now that we have MBL."

"It is, of course, very difficult to tease out the effects of MBL when so many other variables are also being altered. As I said above, I think that MBL helps, but if the students are being told in lecture that all that really matters is solving Halliday/Resnick problems, I think they sometimes just go through the motions with MBL."

B. Comments on Case Studies #4, 5, 6

Consistent with the above case studies, my own experience[22c,41b,76] in conducting field studies of Saturday-morning Socratic Dialogue Inducing (SDI) labs for paid student volunteers enrolled in fairly traditional courses, has been that rather mediocre conceptual development takes place both for the test-group students who take the SDI labs and similar control-group students who do not take the SDI labs. This despite the fact that Table II shows that conceptual development as gauged by FCI pre/post testing (and in some cases MB testing) is much better for SDI-lab containing courses than for traditional courses. The major difference between the field studies and the SDI-containing courses of Table II is that in the former case, attempts are made to graft SDI labs onto traditional courses, whereas in the latter case SDI labs are integrated with IE-type "lectures," IE-type "recitations," and conceptually-oriented exams. The conclusion that, in the field studies, mediocre conceptual development took place for both the test and control group students is based on analyses of (a) videotaped interviews, (b) videotaped lab sessions, and (c) the results of pre/post testing with both FCI and MB exams. Consistent with earlier work,[76] I conclude from the qualitative field-study research[22c,41b,76]; case studies #4-6 above; and the present more quantitative pre/post testing survey (see also ref. 5), that *prominent gain in students' conceptual understanding is much more likely to occur if ALL components of a course are tightly integrated in an IE mode*. Such integration does in fact occur for most of the IE courses of Tables I and II.

Nevertheless, the "better is the enemy of the good." Here "better" would be a completely IE integrated course and "good" might consist of the substitution of an IE component for a traditionally-taught part of a multicomponent traditional course. Departments and schools would be well advised to go for the "good" if organizational problems prohibit the "better," as is frequently the case. The "good" may (a) improve the overall affective and cognitive advancement of students (even though such progress may not show up prominently in FCI gains), (b) serve to educate instructors in IE methods, (c) provide an entry point for the gradual infiltration of more effective pedagogy into mainstream physics education, and (d) initiate a "redesign process"[65] of gradual long-term improvement.

Just before this manuscript was submitted we learned of the encouraging work of Redish *et al.*[34] at the University of Maryland (UM). They have shown that (a) grafting one-hour per week "McDermott Recitation Tutorials"(MRT) [35] onto a traditional calculus-based course for engineering students at UM increased g by about 60% above the control UM traditional course, (b) grafting a one-hour "Targeted MBL Tutorial" (TMT) concentrating on Newton's Third Law (N3) onto a (traditional + MRT) course yielded much higher g's for a 4-question N3 subset of the FCI than are achieved by control sections with traditional recitations - the MBL N3 experiments were adapted from Real Time Physics[33c,e]; and (c) the course-averaged normalized gains for interactive engagement and traditional courses at UM are consistent with the results of the present study.

Case #5, suggests that even despite some bolstering of the lectures with more IE-oriented materials, if the course exams do not include a substantial number of questions or problems which test for the effectiveness of the IE components of a course, then students may have little motivation to take the IE components seriously or appreciate their relevance to scientific thinking and conceptual understanding. In my judgment, in addition to the integration of all components of the course, (a) all instructors in the course, as well as the course syllabus, should clearly indicate to students the goals and methods of science and the importance of IE methods to the students' learning (see e.g., Chap. 1 of ref. 62 and "Objectives of the P201 Course" [22d]), and (b) a substantial fraction of the exam questions should probe the degree of conceptual understanding induced by the IE methods.

Thornton and Sokoloff[33e] have discussed pre/post test data using their Force and Motion Conceptual Evaluation (FMCE) for classes at Tufts University and the University of Oregon. Their "Fig. 7" appears to demonstrate the effectiveness of "Real Time Physics" (RTP) labs when used in otherwise traditional courses. A possible reason for the apparent difference in the FCI results for cases #4-6 and the FMCE results of ref. 33e is simply the difference in the two tests. As indicated above, Paul D'Alessandris has used both exams for three years. He speculates[13a,b] that "........the students view the FMCE as a 'physics' exam; it has lots of graphs and diagrams and is very similar to the homework in RTP. FCI is often viewed as containing questions about reality; balls dropped from buildings, golfballs flying, etc. I have had students (and not just a few) ask me if they are supposed to use formulas on the FCI or just give answers that they think are correct....(cf, "Professor Mazur, how should I answer these (FCI) questions? According to what you taught us, or by the way I think about these things"[19c])...... I believe the FCI may be better than the FCME in indicating what the students really think. The FMCE assesses whether the students have correctly conceptualized Newtonian physics, the FCI tests whether they realize that the world outside of the classroom is Newtonian. In my experience, students can 'understand' Newtonian physics but not believe that the world is actually Newtonian. I think MBL is successful in helping students understand the relationships between force, velocity, and acceleration, but its effect beyond that is unclear to me. (Of course, I don't think the FCI really tests student beliefs as well as some would have us believe.)"

In Table II, the methods employed in case studies #2 - #6 above are indicated by "•?" to indicate the presence of implementation problems.

The present survey shows, in agreement with the preliminary results,[4a,b] that relatively effective methods need not be high tech and need not depend upon Microcomputer Based Laboratories (despite seemingly widespread opinion to the contrary). For example:
(a) three of Paul D'Alessandris's early courses M-PD92a, M-PD92b, and M-PD93 (Table IIb) attained, respectively $g = 0.55$, 0.61, and 0.58 without the use of MBL, as previously mentioned above; (b) an early modeling course ASU-HH-C (Table IIc) achieved $g = 0.52$ without MBL;
(c) my own early SDI courses (IUpre93 – Table IIc) attained a student-averaged $g = 0.53$ without MBL; (d) Eric Mazur's courses EM91-C through EM95-C obtain $g = 0.48$, 0.53, 0.59, 0.64 without MBL. And Concept Tests,[19a,36] Collaborative Peer Instruction in lectures,[17d,32,36] and interactive lectures[17d,19;32] do not require high-tech systems such as Classtalk.[19d,57] As shown in Table II, Concept Tests have been given at Indiana for the past 5 years. These were scored using optical scanning sheets.[36]

IV. SURVEY RATIONALE AND SUGGESTIONS FOR SURVEY IMPROVEMENT

According to Pride *et al.*[85] "The results...(of ref. 85)...demonstrate that responses to multiple-choice questions often do not give an accurate indication of the level of understanding and that questions that require students to explain their reasoning are necessary.... Good performance on a multiple choice test may be a necessary condition, but it is not a sufficient criterion for making this judgment...(of functional understanding of the material)....broad assessment instruments are not sensitive to fine structure and thus may not accurately reveal the extent of student learning. Moreover, *such information does not contribute to a research base that is useful for the design of instructional materials*. " (Our *italics.*)

If it is true that broad assessment instruments such as the FCI/MD and MB are not useful for the design of instructional material but only for increasing "faculty awareness of the failure of many students to distinguish between Newtonian concepts and erroneous common sense beliefs, both before and after instruction in physics,"[85] then the value of surveys such as this one is rather limited. I think that most physics-education researchers would agree that Multiple Choice (MC) tests, even those as carefully crafted as the MD/FCI and MB cannot probe students' conceptual understanding as deeply as can the searching (and labor intensive) analyses of (a) student interviews conducted by physics experts, or, arguably, (b) well-designed, free-response problem exams. In my opinion, MC tests, interviews, problem exams, and case studies all have their advantages, disadvantages, and trade-offs and should be used in combination so as to be mutually supportive whenever possible. The FCI/MD questions, answers, and distractors, were, in fact, developed from extensive interview data.[1a,b;2] The present survey, in addition to the MD/FCI and MB test results, gathers information from detailed questionnaire[4c] responses of instructors, and invokes supplementary case studies (e-mail and telephone interviews) in situations where questionable or unexpected test results were initially obtained. The advantage of carefully designed MC tests (especially if supplemented with other research and testing procedures), is that they allow a standardized measurement with uniform grading over a large population and thus may afford a more practical route to evaluating the effectiveness of methods used in large-enrollment introductory courses at one or many institutions than, by themselves, individual interviews, individually graded exams, or case studies.

The present survey (a) strongly suggests that classroom use of IE methods [i.e., those designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors] can increase mechanics course effectiveness in both conceptual understanding and problem-solving well beyond that achieved with T methods; (b) shows that, for the survey courses, current IE methods fail to produce normalized gains in the High-g region, suggesting the need for improvement of IE strategies in content and/or implementation; (c) gives references to the surveyed IE methods, materials, instructors, and institutions; (d) discusses the various implementation problems that appear to have occurred; and (e) suggests ways to overcome those problems. In my opinion, the foregoing information and suggestions *are* of potential value in designing instructional materials, e.g., current materials need to be improved, new materials should be designed to promote interactive engagement while avoiding the survey-indicated implementation pitfalls. Therefore, I disagree with Pride *et al.* that broad assessment instruments do not "contribute to a research base that is useful for the design of instructional materials."

As discussed in ref. 5a, in my view, the present survey is a step in the right direction but improvements in future assessments might be achieved through (in approximate order of ease of implementation) (1) standardization of test-administration practices; (2) use of a survey questionnaire[4c] refined and sharpened in light of the present experience; (3) more widespread use of standardized tests by individual instructors so as to monitor the learning of their students; (4) use of questionnaires which assess student views on science and learning[73]; (5) observation and analysis of classroom activities by independent evaluators; (5) solicitation of anonymous information from a large random sample of physics teachers; (7) development and use of new and improved versions of the FCI and MB tests, treated with the confidentiality of the MCAT, (8) use of E&M concept tests; and (9) reduction of possible teaching-to-the-test influence by drawing test questions from pools such that the specific questions are unknown to the instructor.[45b]

V. RESEARCH QUESTIONS

Seven research questions raised by the present study and calling for further experimental investigation are listed below.

A. Why Do Some IE Courses Achieve (<g>) < 0.3 , While Others Achieve (<g>) ≥ 0.6?

In Sec. III, I argued that certain implementation problems may be responsible for the placement of some IE courses in the low (<g>) < 0.3 range. In Sec. II-C, I indicated that at the present stage of pedagogical understanding, the particular method or materials used by an instructor may be less important than his/her skill in promoting effective interactive engagement of students. To shed greater light in this area, more thorough case studies (e.g., site visits, videotape analysis of classroom practice, interviews of instructors and students, examination of course material) for courses attaining (<g>) < 0.3 and (<g>) ≥ 0.6 would be of value.

B. Why Are Current IE Methods Relatively Effective For Some Students and Ineffective for Others?

There is commonly a large spread in g's for *individual students* in a course,[83,86] with g's ranging from the maximum g = 1.0 to g = 0.0 (or even negative). Why are current IE methods relatively effective for some student and ineffective for others? To help answer these questions it would be useful to carry out, for any given course, in-depth studies of students in the lower-g < 0.3 and higher-g ≥ 0.6 ranges: e.g., (a) GPA's and SAT's, (b) educational backgrounds, (c) evaluations by teachers, (d) interviews by physics-education researchers, (e) study habits,[87] (f) views on science and learning,[73] (g) attitudes towards the course,[59] and (h) math skills.[1a,16,20a,21,78,84c]

C. Why Do No Survey Courses Achieve (<g>) ≥ 0.7 ?

Jerome Epstein[84a] has suggested that many students entering introductory physics courses may be at cognitive levels too low to benefit from current IE methods, and that this might account for the failure of survey courses to break through the "<g> = 0.7 barrier." It is also possible that deficient cognitive development of entering students contributed to the low-g's of seven of the IE courses (Sec. III). Consistent with the observations of Arons,[43d] Epstein[84b] states: "In large numbers our students... [at Bloomfield College (NJ) and Lehman College(CUNY)]... cannot order a set of fractions and decimals and cannot place them on a number line. Many do not comprehend division by a fraction and have no concrete comprehension of the process of division itself. Reading rulers where there are other than 10 subdivisions, basic operational meaning of area and volume, are pervasive difficulties. Most cannot deal with proportional reasoning nor any sort of problem that has to be translated from English. Our diagnostic test [84c] which has been given now at more than a dozen institutions ...(including Wellesley!)...shows that there are such students everywhere." Epstein and Kolidy have devised and conducted "Freshman Core Programs"[84d] (FCP's) which have substantially increased students' cognitive levels as measured by pre/post testing with standardized reasoning exams. It would be useful to see if (a) individual student scores on Epstein's Diagnostic (ED) correlated with individual-student FCI normalized gains g in single IE courses, (b) average scores on the ED correlated with average normalized FCI gains <g> for many IE courses, and (c) whether or not pre-physics-course FCP's (or similar courses) can raise <g>'s in IE courses.

D. Are There T Courses With Normalized Gains Similar to Those of IE Courses?

A referee has pointed out that although the present study constitutes an "existence proof" that IE courses *can* yield "Medium-g" normalized gains on the FCI, a similar proof for T courses might also be found, thus negating to some extent the conclusion that IE methods are more effective than T methods. That a statistically significant set of n traditional passive-student courses could yield $<<g>>_{nT} \approx <<g>>_{48IE} = 0.48$ seems very unlikely for the following reasons: (a) traditional courses taught by popular and well-regarded teachers have achieved low $<g>_T \leq 0.30$ both at a large state research university[1a] and an ivy-league college,[19] (b) over the past few years results of FCI testing have become fairly well known among physics teachers and even in some research universities, but no normalized gains much above 0.30 have ever, to my knowledge, been reported for traditional courses, (c) that large gains in the conceptual understanding of mechanics could be achieved, on average, by students subjected to passive-student lectures, recipe labs, and algorithmic-problem exams would run counter to two decades of physics-education research.[43] Nevertheless, it may be worthwhile to institute a systematic search for Medium-g (or High-g) traditional courses.

Robert Ehrlich[88] has, in fact, already taken the first steps in this direction. He has pointed out that the "...the size of the sample...(14 courses).... Hake used for the traditional courses was fairly small, so a statistical fluctuation was always a possibility." Seeking to test his conjecture, he promoted pre/post FCI testing in 12 more-or-less traditional courses taught by instructors with whom he was acquainted. These yielded $<<g>>_{12T} = 0.20 \pm 0.06$sd, consistent with the present results $<<g>>_{14T} = 0.23 \pm 0.04$sd. Ehrlich then sought to test the idea that <g> for T courses could be raised simply by including conceptual questions of the type found on the FCI test, both as homework and also test questions. Although he did not obtain enough cooperation to carry out this potentially valuable experiment, it would constitute a worthwhile future research project.

E. Can Courses Taught by Mainstream Teachers Achieve $<g> \geq 0.3$?

As indicated in Sec. IIB, the instructors of this survey were, for the most part, active contributors to the physics-education literature. It is encouraging that high-school courses taught by the participants of Modeling workshops have achieved <g>'s equal to and even exceeding those of this survey.[89] It would be interesting to obtain more FCI and MB data for courses conducted by mainstream teachers who use IE methods but do not normally attend teachers' meetings or publish in the physics-education journals.

F. What is the Relationship of FCI and FMCE Test Results?

Case study #6 discusses two courses at Monroe Community College which incorporated Real Time Physics in the labs, but otherwise traditional pedagogy. Both these courses achieved <g> = 0.25 on the FCI but scores above 65% on the FMCE. Paul D'Alessandris speculates on reasons for the difference, but more systematic and extensive comparison of the results of these two tests should be undertaken before legitimate conclusions can be drawn.

G. Can Grafting of IE Laboratories Onto Traditional Courses Markedly Increase Conceptual Understanding?

Case studies #4-6 suggest that the grafting of Real Time Physics (RTP) labs onto traditional courses at Monroe Community College (MCC) and the University of Louisville (UL) did not markedly increase conceptual understanding as measured by the FCI. On the other hand, use of RTP labs with traditional instruction at Oregon and Tufts drastically increased conceptual understanding as measured by the FMCE.[33e] The apparent discrepancy could be due to (a) difference in the meaning of FCI and FMCE test results as discussed above, (b) more effective implementation of RTP at Oregon/Tufts than at MCC/UL , (c) more effective "traditional" instruction at Oregon/Tufts than at MCC/UL. More research seems to be required before meaningful conclusions can be drawn.

VI. CONCLUSIONS

The present article yields the following answers to the three questions posed in the introduction:

A1. For the present 6542-student survey the most widely used interactive engagement (IE) methods are Collaborative Peer Instruction, 4458 (*all* IE-course students); Microcomputer Based Laboratories, 2704; Concept Tests, 2479; Socratic Dialogue Inducing Labs, 1705; Overview Case Study and Active Learning Problem Sets,1101; Modeling, 885; and research-based text or no text, 660. In addition many other IE methods are being employed. The IE methods are (a) well documented in the literature, (b) can be melded together to enhance one another's strengths, (c) can be modified to suit local conditions, (d) are often available in electronic form, (e) usually offer materials for their implementation, (f) are used in many different types of institutions for diverse student groups by instructors who are usually active contributors to the physics- education literature.

A2. The use of IE methods appears to be necessary but not sufficient for marked improvement over traditional methods as demonstrated by seven courses (N = 717) which utilized IE strategies but achieved <g>'s ranging from 0.21 to 0.26. Case studies suggest that these relatively low average normalized gains were due to difficulties in the implementation and that such problems might be mitigated by (a) apprenticeship education of instructors new to IE methods (Cases 1, 3);
(b) emphasis on the nature of science and learning throughout the course (Case 3); (c) careful attention to motivational factors and the provision of grade incentives for taking IE activities seriously(Case 2); (d) recognition of and positive intervention for potential low-gain students (Case 3); (e) administration of exams in which a substantial number of the questions probe the degree of conceptual understanding induced by the IE methods (Cases 4 – 6); (f) use of IE methods in *all* components of a course and tight integration of those components (Cases 4 – 6). Other suggestions for course improvement gleaned from this survey have been listed in ref. 5a.

A3. The present study gives rise to seven research questions (Sec. V) calling for further experimental investigation.

Epilogue

I am deeply convinced that a statistically significant improvement would occur if more of us learned to listen to our students....By listening to what they say in answer to to carefully phrased, leading questions, we can begin to understand what does and does not happen in their minds, anticipate the hurdles they encounter, and provide the kind of help needed to master a concept or line of reasoning without simply "telling them the answer."....Nothing is more ineffectually arrogant than the widely found teacher attitude that 'all you have to do is say it my way, and no one within hearing can fail to understand it.'....Were more of us willing to relearn our physics by the dialog and listening process I have described, we would see a discontinuous upward shift in the quality of physics teaching. I am satisfied that this is fully within the competence of our colleagues; the question is one of humility and desire.

Arnold Arons, Am. J. Phys. **42**, 157 (1974)

ACKNOWLEDGMENTS

This work received partial support from NSF Grant DUE/MDR9253965. My deepest gratitude goes to those teachers who supplied the invaluable advice, manuscript suggestions, and unpublished data which made this report possible: Albert Altman, Dewayne Beery, Les Bland, Don Boys, Ben Brabson, Bernadette Clemens-Walatka, Paul D'Alessandris, Randall Knight, Priscilla Laws, Cherie Lehman, Eric Mazur, Roger Mills, Robert Morse, Piet Molenaar, Tom O'Kuma, Gregg Swackhamer, Lou Turner, Alan Van Heuvelen, Rick Van Kooten, Mojtaba Vazari, William Warren, and Paul Zitzewitz. We have benefited from additional comments on the manuscript by Amit Bhattacharyya, Ernie Behringer, Sister Marie Cooper, Steve Gottlieb, Ibrahim Halloun, John Hardie, David Hammer, Charles Hanna, David Hestenes, Don Lichtenberg, Tim Long, Joe Redish, Rudy Sirochman, Steve Spicklemire, Richard Swartz, Jack Uretsky, and Ray Wakeland. Intensive e-mail discussions with and among Priscilla Laws, Paul D'Alessandris, Roger Mills, and myself, initially stimulated by the former, were vital to the treatment of implementation difficulties in Sec. III. I thank two JPER referees for valuable comments which greatly improved the manuscript. Socratic Dialogue Labs could not have been successful without many years of assistance and advice from Randall Bird; Willson Hammond; Steve Hovious; Fred Lurie; Jim Sowinski; Ray Wakeland; many post-doctoral-, graduate-, undergraduate-student-instructors; and the 1263 Physics P201 students who partook in SDI labs. All of these SDI-lab contributors helped to demonstrate the classroom effectiveness of the methods long advocated by Arons and now advised by most physics-education researchers. This work would never have been completed without the encouragement and counsel of Arnold Arons, David Hestenes, William Kelly, and Ray Hannapel.

a) Electronic mail: <hake@ix.netcom.com>

References

1. (a) I. Halloun and D. Hestenes, Arizona State University, "The initial knowledge state of college physics students," Am. J. Phys. **53**, 1043-1055 (1985); corrections to the Mechanics Diagnostic (MD) test are given in ref. 16; (b) "Common sense concepts about motion," *ibid.* **53**, 1056-1065 (1985).

2. (a) D. Hestenes, M. Wells, and G. Swackhamer, Arizona State University, "Force Concept Inventory," Phys. Teach. **30**, 141-158 (1992). The FCI is very similar to the earlier Mechanics Diagnostic test and pre/post results using the former are very similar to those using the latter. (b) I. Halloun, R.R. Hake, E.P. Mosca, and D. Hestenes, Force Concept Inventory (Revised, 1995) in ref. 19b and password protected at <http://modeling.la.asu.edu/modeling.html>. Comparisons of gains attained with the revised FCI on courses with a long history of FCI pre/post testing at Harvard and Indiana University suggest that pretest averages may tend to be somewhat lower with the revised FCI (see courses EM-95C and IU95F of Table I), but that average normalized gain <g> values are not much affected. (c) Gregg Swackhamer, Glenbrook North High School (public), private communication, 4/96. (d) D. Hestenes, "Guest Comment: Who needs physics education research!?" Am. J. Phys. **66**, 465-467 (1998).

3. D. Hestenes and M. Wells, Arizona State University "A Mechanics Baseline Test," Phys. Teach. **30**, 159-166 (1992). The test is also in ref. 19b and and password protected at <http://modeling.la.asu.edu/modeling.html>.

4. R.R. Hake, (a) "Assessment of Introductory Mechanics Instruction," AAPT Announcer **23**(4), 40 (1994); (b) "Survey of Test Data for Introductory Mechanics Courses," *ibid.* **24**(2), 55 (1994); (c) "Mechanics Test Data Survey Form," at <http://carini.physics.indiana.edu/SDI/>.

5. R.R. Hake (a) "Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. **66**, 64 -74 (1998); (b) "Evaluating conceptual gains in mechanics: A six-thousand student survey of test data," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE*, ed. by E.F. Redish and J.S. Rigden, (AIP, Woodbury, NY, 1997). p. 595 - 604; in that paper the 7 Low-g IE courses, deemed to have implementation problems as evidenced by instructors' comments, were omitted from the IE averaging so as to obtain $<<g>>_{41IE} = 0.52 \pm 0.10$sd. I now think that the present treatment is preferable.

6. Bernadette Clemens-Walatka, Sycamore High School, private communications, 6/95, 4/96, 11/96; (a) "An Interdisciplinary Option, Coordinating Pre-calculus, First Year High School Physics, and Computer Technology," AAPT Announcer **24**(2), 53 (1994); "Physics and Precalculus: Common Connections in a Technology Rich Environment," *ibid.* **26**(2), 88 (1996); (b) a first-year college prep course. For the work in "a" BC and her math counterpart were awarded the prestigious $12,000 GIFT (Growth Initiatives For Teachers) grant sponsored by GTE for integrating science and mathematics at the secondary level. For "a" and for the development of her own original MBL labs, BC received the $2500 Tandy award at a recent National Science Teachers Association convention.

7. Cherie Lehman, West Lafayette High School, (a) private communications, 1/94, 3/96. (b) See also C. Lehman, "Investigating Motion with the CBL Motion and Force Probes," AAPT Announcer **25**(2), 47 (1995); "Modeling an Exponential Decay with the CBL," *ibid.* **26**(2), 68 (1996).

8. A.L. Ellermeijer, B. Landheer, P.P.M. Molenaar, "Teaching Mechanics through Interactive Video and a Microcomputer-Based Laboratory," 1992 NATO Amsterdam Conference on Computers in Education, Springer Verlag, in press; private communications from P.P.M. Molenaar, 6/94, 4/96.

9. (a) Lou Turner, Western Reserve Academy, private communications, 1/94, 6/94, 6/95, 3/96. Western Reserve Academy is selective private coeducational prep school; (b) private communication 3/26; (c) "Using Air as an Analogy to Understand Electricity," AAPT Announcer **25**(4), 66 (1995).

10. Robert Morse, St. Albans School, (a) private communications, 11/94, 6/95, 3/96, 11/96. The average posttest score for Morse’s students was probably lowered by the fact that at his school "seniors with a grade of B or better....(28% of his students in RM94 and 12% in RM95).... are exempt from the final exam," and thus did not take the FCI posttest. Saint Albans is a selective boy’s prep school. (b) See also R.A. Morse, "Acceleration and Net Force: An Experiment with the Force Probe," Phys. Teach. **31**, 224-226 (1993); (c) "The Classic Method of Mrs. Socrates," *ibid.* **32**, 276-277 (1994), but see ref. 39.

11. Dewayne Beery, Buffalo State College, private communications, 4/94, 5/94, 3/96. Minority enrollment in this course for science majors was about 25%. Buffalo State is an "urban public school which is not very selective."

12. William Warren, Lord Fairfax Community College, private communication, 3/96, 5/96.

13. Paul D’Alessandris, Monroe Community College (MCC), (a) private communications, 11/94, 5/95, 3/96, 4/96. These data are for a "lecture-based curriculum" taught by others. (b) private communication, 11/94, 5/95, 3/96, 4/96, 11/96. These data are for an "active-learning curriculum." No grade incentives for performance on the posttest FCI are given at MCC. These might raise the FCI gains and thus g by about 5%. (c) See also P. D’Alessandris, "The Development of Conceptual Understanding and Problem-Solving Skills through Multiple Representations and Goal-less Problems," AAPT Announcer **24**(4), 47 (1994); (d) "Addressing Alternative Concepts in Rotational Motion through Microcomputer Based Laboratories and *Interactive Physics*," *ibid.* **25**(4), 49 (1995); "Assessment of a Research-Based Introductory Physics Curriculum," *ibid.* **25**(4), 77 (1995); "Repercussions of an NSF-ILI Grant on Monroe Community College," *ibid.* **26**(2), 46 (1996); (e) SPIRAL Physics workbooks are available at <FTP@eckert.acadcomp.monroecc.edu>; (f) "SPIRAL Physics active learning workbooks," preprint, 4/96. MCC is an open admissions two-year college which "draws a very mixed bag of students; urban poor, suburban underachievers, and rural everything else. The student body is, if anything, middle to lower-middle class."

14. Priscilla Laws, Dickinson College, (a) private communication to D. Hestenes, 1992; (b) private communications, 5/95, 4/96. (c) See also "Workshop physics: Replacing lectures with real experience" in *Proc. Conf. Computers in Physics Instruction*, ed. by E. Redish and J. Risley (Addison-Wesley, 1989), pp. 22-32; "Calculus-Based Physics Without Lectures," Phys. Today **44**(12), 24-31 (1991); "Millikan Lecture 1996: Promoting active learning based on physics education research in introductory physics courses," Am. J. Phys. **65**, 13-21 (1997). H. Pfister and P. Laws, "Kinesthesia-1: Apparatus to Experience 1-D Motion," Phys. Teach. **33**, 214-220 (1995). (d) For a case study of Workshop Physics see ref. 82a, p. 103-107. (e) P. Laws *et al.*, *Workshop Physics Activity Guide* (Wiley, 1997). See also at <http://physics.dickinson.edu/> and <http://galileo/harvard.edu/> under "Hands On Methods." Dickinson is a selective 4-year college.

15. Thomas O’Kuma, Lee College, (a) private communication, 5/95, 4/96. (b) See also C.J. Hieggelke and T.L. O’Kuma, "MBL Rotational Motion and Magnet Field Labs," AAPT Announcer **25**(2), 62 (1995); T.L. O’Kuma, "More CE/OCS/ MBL Results in the Introductory Physics Course," *ibid*. **25**(2), 96 (1995); C. J. Hieggelke, T. O’Kuma, and D. Maloney, "Ranking Tasks," *ibid.* **25**(4), 77 (1995); T.L. O’Kuma, C.J. Hieggelke, and A. Van Heuvelen, "Using Bar Charts in Introductory Physics," *ibid.* **25**(4), 78 (1995); R.B. Clark and T.L. O’Kuma, "Two-Year College Physics Faculty Enhancement Program (PEPTYC)," *ibid.* **26**(2), 67 (1996). Lee is an open admissions 2-year college with a majority of students from low to low middle income families. It has over 30% minorities and over 56% women students. The average student age is 29. According to O’Kuma, Lee is fairly typical of most two-year community colleges.

16. I. Halloun and D. Hestenes, Arizona State University, "Modeling instruction in mechanics," Am. J. Phys. **55**, 455-462 (1987). The data shown in Table 1c are for "Test Group #3," for which the Modeling method was most fully implemented.

17. Alan Van Heuvelen, Ohio State University, (a) private communication, 8/94. The Physics 105 courses was taught at Arizona State University and was composed of "academically deprived students." (b) Private communication, 4/96, 11/96, regarding courses at Ohio State. No grade incentives for performance on the posttest FCI are given at Ohio State. These might raise the FCI gains and thus g by about 5%. (c) See also, A. Van Heuvelen, "Learning to think like a physicist: A review of research-based instructional strategies," Am. J. Phys. **59**, 891-897 (1991); (d) "Overview, Case Study Physics," *ibid*., 898-907 (1991); (e) "Experiment Problems for Mechanics," Phys. Teach. **33**, 176-180 (1995); (f) "ActivPhysics" CD-ROM with workbook is available from Addison Wesley Interactive, <http:awi.aw.com/products.html#ActivPhysics>. (g) For a case study of Van Heuvelen's methods see ref. 82a, p. 100 - 103. Some materials are available commercially from Hayden-McNeil Publishing Inc., 47461 Clipper St. Plymouth, MI 48170; 313-455-7900.

18. Randall Knight, California Polytechnic State University (San Luis Obispo), private communications regarding Cal Poly courses taught by (a) others, 4/94; (b) himself, 4/94, 3/96, 11/96. (c) R.D. Knight, *Physics: A Contemporary Perspective* (Addison-Wesley-Longman, 1997); (d) "The Vector Knowledge of Beginning Physics Students," Phys. Teach. **33**, 74-78 (1995).

19. Eric Mazur, Harvard University, (a) private communications, 5/95, 4/96, 11/96; a course for science (but not physics) majors. Although there *is* a small grade incentive for performance on the FCI, students punch in answers to both the pre- and post-test FCI electronically in such a way that previous answers cannot be changed. Since this handicap might negatively affect the posttest (grade incentive) averages more than the pretest (no grade incentive) averages, it's possible that Harvard g's are artificially lowered by a few percent. (b) See also E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, 1997), contains the 1995 revision of the FCI; (c) "Qualitative vs. Quantitative Thinking: Are We Teaching the Right Thing?" Optics and Photonics News **3**, 38 (1992). (d) For assessment data, course syllabus, *User's Manual*, information on Classtalk, and examples of Concept Tests see at <<http://galileo.harvard.edu/>; (e) "Are science lectures a relic of the past?" Physics World **9**(9), 13-14 (1996). (f) For a case study of Mazur's methods, see ref. 82a, Chap. 8, p. 114-122, "Students Teaching Students, Harvard Revisited."

20. R.R. Hake, Indiana University, (a) "Promoting student crossover to the Newtonian world," Am J. Phys. **55**, 878-884 (1987); (b) "My Conversion To The Arons-Advocated Method Of Science Education," Teaching Education **3**(2), 109-111 (1991).

21. S. Tobias and R.R. Hake, Indiana University, "Professors as physics students: What can they teach us?" Am. J. Phys. **56**, 786-794 (1988).

22. R.R. Hake, Indiana University, (a) unpublished data for Physics P201, a course for science (but not physics) majors, enrolling primarily pre-meds and pre-health-professionals. At Indiana, the FCI posttest normally counts for half the final-exam grade (about 12% of the final-course grade). (b) "Socratic Pedagogy in the Introductory Physics Lab," Phys. Teach. **30**, 546-552 (1992) and at <http://carini.physics.indiana.edu/SDI/>; (c) "Socratic Dialogue Labs in Introductory Physics," in *Proceedings of the 1995 Conference on New Trends in Physics Teaching*, ed. by J. Slisko (Univ. of Puebla; Puebla, Mexico, in press). (d) For a summary of recent work and a updated listing of electronically available SDI materials (e.g., manuals, teacher's guides, sample lab exams, equipment set-up lists) see <http://carini.physics.indiana.edu/SDI/> and <http://galileo.harvard.edu/> under "Hands On Methods," or contact R. R. Hake at <hake@ix.netcom.com>. Ref. 22b and SDI Labs #1-3 (versions of 10/93) are available on the Fuller-Zollman CD-ROM InfoMall, 3rd field-test and commercial versions (ref. 51).

23. R. Van Kooten, R.R. Hake, F.M. Lurie, and L.C. Bland, Indiana University, unpublished data. Team taught Physics P201 course (science, but not physics, majors) in Spring 1995.

24. L.C. Bland, B.B. Brabson, R.R.Hake, J.G. Hardie, and E. Goff, Indiana University, unpublished data. Team taught P201 course in Fall 1995.

25. P. W. Zitzewitz, Univ. of Michigan (Dearborn), private communication, 12/95, 3/96. See also P.W. Zitzewitz, "Evaluation of General Physics (Mechanics) Laboratories that Include 'Experiment Problems'," AAPT Announcer **26**(2), 94 (1996).

26. (a) Don Boys, Univ. of Michigan (Flint), private communication, 8/94, 3/96. (b) See also M. Vaziri and D. Boys "Improving the Introductory Mechanics Course," AAPT Announcer **24**(2), 81 (1994).

27. Roger Mills, Univ. of Louisville, private communication (1/95) regarding a course taught by others in the Fall of 1994.

28. Roger Mills, Univ. of Louisville, (a) private communications (4/95, 3/96, 4/96, 5/96) regarding a course in Spring 1995 in which RM managed the labs, while the lectures and recitations were taught by others in a rather traditional manner; (b) private communication, D.R. Sokoloff to R. Mills; (c) private communication (3/96, 4/96, 5/96) regarding a course in Summer 1995 in which RM managed the labs and also met the lectures. See also, R. E. Mills, "Introductory Physics Recitations and Laboratory Improvement," AAPT Announcer **26**(2), 47 (1996).

29. Albert Altman, Univ. of Massachusetts at Lowell, private communications, 5/95, 4/96. Enrollment of minorities and non-native English speakers was relatively high; see also at the WWW address given in ref. 19d.

30. D.W. Johnson, R.T. Johnson, and K.A. Smith, *Cooperative Learning: Increasing College Faculty Instructional Productivity* (George Washington University, 1991).

31. P. Heller, R. Keith, S. Anderson, "Teaching problem solving through cooperative grouping, Part 1: Group vs individual problem solving," Am. J. Phys. **60**, 627-636 (1992); P. Heller and M. Hollabaugh "Teaching problem solving through cooperative grouping, Part 2: Designing problems and structuring groups," *ibid*., p. 637-644. See <http://www.physics.umn.edu/groups/physed/index.html>.

32. Collaborative Peer Instruction methods quite similar to Mazur's have been recently been described by D.E. Meltzer and K. Manivannan, "Promoting Interactivity in Physics Lecture Classes," Phys. Teach. **34**, 72-76 (1996). Recent work on interactive-engagement methods for large-enrollment "lecture" classes is reviewed. See also J. Poulis, C. Massen, E. Robens, and M. Gilbert, "Physics lecturing with audience paced feedback," Am. J. Phys. **66**, 439-441 (1998).

33. (a) R.F. Tinker, "Computer Based Tools: Rhyme and Reason," in *Proc. Conf. Computers in Physics Instruction*, ed. by E. Redish and J. Risley (Addison-Wesley, 1989*)*, pp. 159-168; (b) R. K. Thornton, "Tools for scientific thinking: Learning physical concepts with real-time laboratory measurement tools," *ibid.* pp. 177-189; "Tools for scientific thinking - microcomputer-based laboratories for physics teaching," Phys. Educ. **22**, 230-238 (1987); R.K. Thornton and D. R. Sokoloff, "Learning motion concepts using real-time microcomputer-based laboratory tools," Am. J. Phys. **58**, 858-867 (1990); R.K. Thornton, "Conceptual dynamics: changing student views of force and motion," in *Thinking Physics for Teaching,* ed. by C. Bernardini, C. Tarsitani, and M. Vicintini (Plenum, 1995); (c) D.R. Sokoloff, P.W. Laws, and R.K. Thornton, "Real Time Physics, A New Interactive Introductory Lab Program," AAPT Announcer **25**(4), 37 (1995). (d) For commercial MBL equipment see L.K. Wilkinson, J. Barnes, M.H. Gjertsen, and J.S. Risley "A Buyer's Guide to Microcomputer-Based Laboratory Equipment in Physics Education," Computers in Physics **9**, 185-199 (1995). (e) R.K. Thornton and D.R. Sokoloff, "Assessing student learning of Newton's laws : The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula, Am. J. Phys. **66**, 338-351 (1998); the FMCE test can be downloaded by password holders at <http://physics.dickinson.edu/PersonalPages/PLaws/WWPages/InstRes/ConceptExams/ConceptExamsDL>; See also at <http://www.tufts.edu/as/csmt/research.html>. (f) D.R. Sokoloff and R.K. Thornton, "Using Interactive Lecture Demonstrations to Create an Active Learning Environment, Phys. Teach **35**, 340-347 (1998). (g) R.K. Thornton and D. R. Sokoloff, "Real Time Physics: Active Learning Laboratory," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE*, ed. by E.F. Redish and J.S. Rigden, (AIP, Woodbury, NY, 1997), p. 1101-1118.

34. E.F. Redish, J.M. Saul, and R.N. Steinberg, "On the effectiveness of active-engagement microcomputer-based laboratories," Am. J. Phys. **65**, 45-54 (1997). See also at <http://physics.umd.edu/rgroups/ripe/perg/research.html>.

35. L.C. McDermott, P.S. Shaffer, and M.D. Somers, "Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine," Am. J. Phys. **62**, 46-55 (1994). L.C. McDermott, S. Vokos, and P.S. Shaffer, "Sample Class on *Tutorials in Introductory Physic*s," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE*, ed. by E.F. Redish and J.S. Rigden, (AIP, Woodbury, NY, 1997). p. 1007 - 1009. McDermott Recitation Tutorials (MRT) require some a modification in the exams and management of a traditional course beyond just the recitation. See <http.www.phys.washington.edu/groups/peg/>; <http://galileo/harvard.edu/> under "Small Group Methods."

36. At Indiana only one Concept Test (CT) per week is given on Monday, each consisting of one conceptual problem with 3 related questions of increasing difficulty. The Monday CT, given near the end of the class, probes student understanding of concepts emphasized during the previous week and during the preceding "lecture." Each student enters his/her individual and collaborative group response onto an optical scanning sheet (picked up on entering the "lecture" hall). Only the collaborative-group (not the single-student) responses count towards the course grade. If possible, the CT is discussed when student interest is at a peak, just after the scanning sheets have been collected.

37. D. Hestenes, (a) "Towards a modeling theory of physics instruction," Am. J. Phys. **55**, 440-454 (1987); (b) "Modeling Games in the Newtonian World," *ibid.* **60**, 732-748 (1992); (c) "Modeling Methodology for Physics Teachers," *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE*, ed. by E.F. Redish and J.S. Rigden, (AIP, Woodbury, NY, 1997), p. 935- 957.

38. M. Wells, D. Hestenes, and G. Swackhamer, "A modeling method for high school physics instruction, Am. J. Phys. **63**, 606-619 (1995). See also at <http://modeling.la.asu.edu/modeling.html>.

39. The Socratic method has been sadly neglected by physics instructors, despite its demonstrated effectiveness. Possible reasons are (1) the competing allure of the quick high-tech fix (rather than slow deep-thought redesign) of science education, (2) the degree of understanding and commitment required of instructors, (3) unfamiliarity with and misunderstanding of the method. That method, employed so successfully (see Table II and refs. 20-22, 34, 35, 43d) is *not* derived from the *classic* Socrates of Plato's *Meno* (cf. ref 10c), but rather from the *historical* Socrates as researched by G. Vlastos, private communication and *Socrates, Ironist and Moral Philosopher* (Cornell Univ. Press, 1991), esp. Chap. 2, "Socrates contra Socrates in Plato."

The essence of Socrates was set forth by Howard Gardner in "The Academic Community Must Not Shun the Debate Over How to Set National Educational Goals," *The Chronicle of Higher Education*, 8 Nov. 1989: "*If Confucius can serve as the Patron Saint of Chinese education, let me propose Socrates as his equivalent in a Western educational context – a Socrates who is never content with the initial superficial response, but is always probing for finer distinctions, clearer examples, a more profound form of knowing. Our concept of knowledge has changed since classical times, but Socrates has provided us with a timeless educational goal – ever deeper understanding.*"

For good discussions of the Socratic method including an attempt to encapsulate it in "production rules" see A. Collins, "Processes in Acquiring Knowledge," in *Schooling and Acquisition of Knowledge*, ed. by R.C. Anderson, R.J. Spiro, and W.E. Montague (Lawrence Erlbaum, 1977); A. Collins and A.L. Stevens, "Goals and strategies for inquiry teachers," in *Advances in Instructional Psychology*, vol. II, ed. by R. Glaser (*ibid.*, 1982); A. Collins and A.L. Stevens, "A cognitive theory of interactive teaching," in *Instructional Design Theories and Models: An Overview*, ed. by C.M. Reigeluth (*ibid.*, 1983). For a neural-network justification of the dialectic method see D. Hestenes in ref. 37a.

Socratic dialogue is not a panacea, but is most useful for finding out what and how students are thinking, guiding them to construct their own understanding of difficult physics concepts, and for conveying scientific approaches and reasoning skills (see Arons in ref. 43d, p. 325). For less difficult instructional tasks, other methods may be more efficient. The complementarity of Socratic, didactic, and coaching instruction is discussed by D. Perkins, *Smart Schools* (Free Press, 1992).

40. R.R. Hake and R. Wakeland, " 'What's F? What's m? What's a?': A Non-Circular SDI-TST-Lab Treatment of Newton's Second Law" in *Conference on the Introductory Physics Course on the occasion of the retirement of Robert Resnick,* Jack Wilson, ed. (Wiley, 1997), p. 277-283. (TST ≡ Tools for Scientific Thinking, see ref. 33b.)

41. (a) A. Bhattacharyya, R.R. Hake, R. Sirochman, "Improving Socratic Dialogue Inducing (SDI) Labs," AAPT Announcer **25**(2), 80 (1995); (b) R.R. Hake and R. Bird, "Why Doesn't The Water Fall Out Of The Bucket? Concept Construction Through Experiment, Discussion, Drawing, Dialogue, Writing, and Animations," *ibid.* **25**(2), 70 (1995); (c) R. Bird and R.R. Hake, "Force Motion Vector Animations on the Power Mac,"*ibid.* **25**(2), 80 (1995). (d) R.R. Hake, "More on Coriolis myths and draining bathtubs," Am. J. Phys. **62**, 1063 (1994) (see SDI Lab #3 Appendix: "Rotating Reference Frames").
(e) A. Roychoudhury, D. Gabel, and R.R. Hake, "Inducing and Measuring Conceptual Change in Introductory-Course Physics Students," AAPT Announcer **19**(4), 64 (1989).

42. (a) J.L. Uretsky, "Using 'Dialogue Labs' in a Community-College Physics Course," Phys. Teach. **31**, 478-481 (1993); (b) N. C. Steph, "Improving the Instructional Laboratory with TST and SDI Labs: Mixing, Matching, and Modifying Ideas," AAPT Announcer **21**(4), 61 (1991). (TST ≡ Tools for Scientific Thinking, see ref. 33b.)

43. For overviews of physics-education research see, e.g.: (a) *Toward a scientific practice of science education*, ed. by M. Gardner, J.G. Greeno, F. Reif, A.H. Schoenfeld, A. diSessa, and E. Stage (Erlbaum, 1990). (b) *Research in Physics Learning: Theoretical Issues and Empirical Studies,* R. Duit, F. Goldberg, and H. Niedderer, eds. (Institute for Science Ed., Kiel, 1992). (c) A. Van Heuvelen, ref. 17c. (d) A. B. Arons, *A Guide To Introductory Physics Teaching* (Wiley, 1990), reprinted with minor updates in *Teaching Introductory Physics* (Wiley, 1997) [also contains *Homework and Test Questions for Introductory Physics Teaching* (Wiley, 1994) along with a new monograph "Introduction to Classical Conservation Laws"]; (e) "Generalizations to be drawn from results of research on teaching and learning" in *Thinking Physics for Teaching,* ed. by C. Bernardini, C. Tarsitani, and M. Vicintini (Plenum, 1995). (f) A. A. diSessa, "Toward an Epistemology of Physics," Cog. and Inst. **10**(2,3), 105-225 (1993). (g) D. Hammer, "More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research," Am. J. Phys. **64**, 1316-1325 (1996); (h) L. C. McDermott and E.R. Redish, Research in physics education resource letter, in preparation. (i) A review of recent physics-education-research based texts is given by J. Amato, Phys. Today **49**(12), 46-51 (1996). (j) L. C. McDermott, "Students' conceptions and problem solving in mechanics" in *Connecting Research in Physics Education with Teacher Education,* A. Tiberghien, E.L. Jossem, and J. Barojas, eds. (I.C.P.E., 1998); also at <http://aapt.org/> under "Physical Science Resource Center"/"21st Century Classrooms."

44. E.F. Redish, "Implications of cognitive studies for teaching physics," Am. J. Phys. **62**, 796-803 (1994); (b) J. Mestre and J. Touger, "Cognitive Research--What's in It for Physics Teachers?" Phys. Teach. **27**, 447-456 (1989).

45. (a) F. Reif, "Millikan Lecture 1994: Understanding and teaching important scientific thought processes," Am. J. Phys. **63**, 17-32 (1995); (b) "Guest Comment: Standards and measurements in physics – Why not in physics education?" *ibid.* **64**, 687-688 (1996). See <http://cil-andrew.cmv.edu/>

46. (a) L. C. McDermott, "A perspective on teacher preparation in physics and other sciences: The need for special science courses for teachers," Am. J. Phys. **58**, 734-742 (1990); (b) "Millikan Lecture 1990: What we teach and what is learned: Closing the gap," *ibid.* **59**, 301-315 (1991); "Guest Comment: How we teach and how students learn: A mismatch?" *ibid.* **61**, 295-298 (1993);

47. R.L. Taylor, "Using the Graphing Calculator - in Physics Labs," Phys. Teach **33**, 312-314 (1995); C. Brueningsen and W. Bower,"Using the Graphing Calculator - in Two Dimensional Motion Plots," *ibid.*, p. 314-316; D.R. Stump, "Using the Graphing Calculator - in Sample Physics Problems," *ibid*, p. 317-318; L.G. Stowe, "Using the Graphing Calculator - in Collision Problems," *ibid*, p. 318-319.

48. For a directory to Physics Academic Software (and other physics education software) see J.S. Risley and M.H. Gjertsen "Tenth Anniversary CIP Educational Software Directory," Computers in Physics **11**, 49-79 (1997) and also <http://www.aip.org/pas/>

49. R.G. Fuller, "Millikan Lecture 1992: Hypermedia and the knowing of physics: Standing upon the shoulders of giants," Am. J. Phys. **61**, 300-304 (1993). D. Zollman, "Millikan Lecture 1995: Do they just sit there? Reflections on helping students learn physics," *ibid.* **64**, 114-119 (1996).

50. (a) D.L Goodstein and R.P. Olenick, "Making 'The Mechanical Universe'," Am. J. Phys. **56**, 779-785 (1988). For information on the Mechanical Universe videotapes see <http://www.projmath.caltech.edu/mu.htm>, (b) Excellent animations depicting the graphical meanings of the derivative and integral, due to James Blinn, are contained this series.

51. See, e.g., R.G. Fuller and D.A. Zollman, "Doing the Physics InfoMall: What the CD-ROM Developers Don't Tell You," AAPT Announcer **25** (4), 75 (1995). See also ref. 49. The InfoMall is now available commercially from The Learning Team, 84 Business Park Suite 307, Armonk, NY 10504, 1-800-793-8326.

52. R. Unruh, L.L. Countryman, and T. Cooney, "The Prisms Approach," Sci. Teach., May 1992, p. 36-41.

53. D. P. Maloney, "Ranking Tasks: A new type of test item,'" J. Coll. Sci. Teach **16**, 510-514 (1987).

54. D.A. Zollman and R.G. Fuller, "Teaching and Learning Physics with Interactive Video," Phys. Today **47**(4), 41-47 (1994). See also ref. 49.

55. L. C. McDermott, P. Shaffer, and M. Rosenquist, *Physics by Inquiry* (Wiley, 1996).

56. A.F. Burr, "*Interactive Physics* 3.0 Delights Window Users Also," Computers in Physics **11**, 275-279 (1997); P. Martinez-Jimenez, E. Casado, J.M. Martinez-Jimenez, M. Cuevas-Rubino, D. Gonzalez-Caballero, and F. Zafra-Lopez, "*Interactive Physics* Simulations Appeal to First-Year Students," Computers in Physics **11**, 31-35 (1997). For an IP Problem Set running on Macintosh computers with IP-II installed, see at <http://info.itp.berkeley.edu/Vol1/Contents.html>.

57. *Classtalk*, a classroom communication system, is a product of Better Education, Inc., 4824 George Washington Highway, Suite 103; Yorktown, VA 23692, 804-898-1897, <info@bedu.com>; J.C. Webb, G.R. Webb, R Caton, and F. Hartline, "Collaborative Learning and Insights on Students' Thinking: Computer Technology Offers New Possibilities for Large Classes," AAPT Announcer **24** (4), 64 (1994); *Classtalk* use at Harvard for ConcepTests and student evaluation of *Classtalk* is shown in ref. 19d; *Classtalk* usage at the Univ. of Massachusetts (Amherst) is discussed at <http://www-perg.phast.umass.edu/pages/MainMenu.html>. A commercial *wireless* classroom communication system is described by R.A. Burnstein and L.M. Lederman, "Report on progress in using a wireless keypad response system," *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE*, ed. by E.F. Redish and J.S. Rigden, (AIP, Woodbury, NY, 1997). p. 531-537. Pedagogical advantages and utilization of Classroom Communication Systems (CCS) are discussed by R.J. Dufresne, W.J. Gerace, W.J. Leonard, J.P. Mestre, and L. Wenk, "*Classtalk*: A classroom communication system for active learning," J. Computing in Higher Ed. **7**(2), 3-47 (1996). CCS may allow a cost-effective semi-Socratic approach (refs. 20-22, 34, 35, 39-42, 43d) to instruction in large-enrollment "lecture" sections.

58. C. Schwartz, unpublished work describing this physicist's invention of "Minute Papers" as described by R.C. Wilson, J. of Higher Ed. **57**(2), 196-211 (1986) and private communication. Minute Papers are discussed (with no mention of Schwartz) by T.A. Angelo and K.P. Cross, *Classroom Assessment Techniques*, (Jossey-Bass, 1993, 2nd ed.) p. 148-153. Following Schwartz, during the last minute of a physics lecture at Indiana, students write down and submit answers to two questions: (1) What is the most significant *scientific* idea that you learned today in lecture? (2) What *physics* question is uppermost in your mind as you leave the lecture? Some advantages are: (a) provides feedback on what the students are (or more usually are not) learning in lecture, (b) supplies a list of student questions to which the lecturer may wish to respond, (c) provides an attendance record (the small grade benefit given for regular presence dramatically increases the lecture attendance), (d) identifies students who may need extra help or enriched material, (e) encourages students to take responsibility for their own learning, (f) motivates students to listen more actively.

59. R.R. Hake and J. C. Swihart, "*DI*agnostic *S*tudent *CO*mputerized *E*valuation of Multicomponent Courses (DISCOE)" Teaching and Learning (Indiana University), January 1979, on the Web at <http://carini.physics.indiana.edu/SDI/>.

60. *First Class* is an electronic bulletin board and conferencing system which is in extensive use for large-enrollment classes at Indiana University.

61. J.A. Dekker, *Motion* (in Dutch) (Univ. of Amsterdam, 1990).

62. F. Reif, *Understanding Basic Mechanics* (Text and Workbook) (Wiley, 1994).

63. The term "Collaborative Peer Instruction" (CPI) is used here in preference to Mazur's "Peer Instruction" (ref. 19b) because "peer *instruction"* does not necessarily require "peer *collaboration"* (e.g., one student in a group could lecture to the others); but most investigators indicate that peer instruction and peer collaboration in physics "lectures," "recitations," and labs almost always occur together in cases where peer instruction is effective (as might be expected from physics education research). "Collaborative Peer Instruction" is preferred over "Cooperative Leaning" or "Collaborative Learning" because the latter terms prejudge the effectiveness of the methods.

64. See, e.g., (a) J.S. Rigden, D.F. Holcomb, and R. DiStefano, "The Introductory University Physics Project," Phys. Today **46**(4), 32-37 (1993); R. DiStefano, "The IUPP Evaluation: What we were trying to learn and how we were trying to learn it," Am. J. Phys. **64,** 49-57 (1996); "Preliminary IUPP results: Student reactions to in-class demonstrations and to presentations of coherent themes," *ibid*., **64,** 58-68 (1996); R. DiStefano and D.F. Holcomb,"The IUPP Evaluation: How did the models fare?" and "The IUPP Evaluation: How did the goals fare?" available from the American Institute of Physics (1-800-809-2247); L.A. Coleman, D.F. Holcomb, and J.S. Rigden, "The Introductory University Physics Project 1987-1995: What has it accomplished? Am. J. Phys. **66,** 124-137 (1998); (b) Modeling Workshop Project, ref. 38 and <http://modeling.la.asu.edu/modeling.html>. (c) R.P. Olenick, "C3P (Comprehensive Conceptual Curriculum for Physics)," AAPT Announcer **26**(2), 68 (1996), <http://phys.udallas.edu>. (d) AAPT, "Powerful Ideas in Science," <http://www.aapt.org>; (e) Constructing Physics Understanding, F. Goldberg, P. Heller, and S. Bendall,<http://cpuproject.sdsu.edu/CPU/>. See also listings at the AAPT's Physical Science Resouce Center, <http://aapt.org/>.

65. K.G. Wilson and B. Daviss, *Redesigning Education* (Henry Holt, 1994) (a goldmine of valuable references), see also at <http://www-physics.mps.ohio-state.edu/~kgw/RE.html>; K.G. Wilson, C.K.Barsky, and B. Daviss, "A Model Development Concept (MDC) for Education," preprint, 1996.

66. Problem and conceptual-question sources due to instructors referenced in Tables I or II are (a) F. Reif, ref. 62; (b) E. Mazur, ref. 19d and <http://galileo.harvard.edu/>; (c) P. D'Alessandris, "Goal-less Problems," ref. 13c,e; (d) Priscilla Laws, Alternative Homework Assignments at <http://physics.dickinson.edu/PersonalPages/PLaws/ABPPages/AHA>; (e) R..D. Knight, ref. 18c; (f) A. Van Heuvelen, "Experiment Problems" and "ActivPhysics" Problems, ref. 17e,f; (g) R.R. Hake, "Out-of-Lab Problems," ref. 22d; (h) P. Heller, "Context-rich problems," ref. 31 and <http://www.physics.umn.edu/groups/physed/index.html>; (i) D. P. Maloney, "Ranking Tasks, ref. 53; "Fill-in problems," J. Coll. Sci. Teach. **12**, 104-107 (1982).

Other useful problem sources are: (j) A.B. Arons, ref. 43d; (k) R.E. Gibbs, *Qualitative Problems for Introductory Physic*s (Kendall Hunt, 1990); (l) C.W. Camp and J.J. Clement, *Preconceptions in Mechanics* (Kendall Hunt, 1994); (m) L.C. Epstein, *Thinking Physics* (Insight Press, 1990); (n) L. Nedelsky, *Science Teaching and Testing* (Harcourt, Brace, and World, 1965); (o) H. R. Crane, "Problems for Introductory Physics," Phys. Teach. **7**, 371-378 (1969); *ibid.* **8,** 182-187 (1970); "Better Teaching with Better Problems and Exams" Phys. Today **22**(3), 134-135 (1969); (p) F. J. Blatt, *Principles of Physics* (Allyn and Bacon, 3rd ed., 1989), excellent conceptually-oriented multiple-choice questions at the end of each chapter; (q) C. E. Swartz, *Used Math* (AAPT, 1993); C.E. Swartz and T. Miner, *Teaching Introductory Physics: A Sourcebook* (AIP Press, 1997), chap. 1; (r) Univ. of Maryland, Research in Physics Education: "Estimation (Fermi) Problems" at <http://physics.umd.edu/rgroups/ripe/perg/fermi.html>; "Thinking Problems in Mechanics" at <http://physics.umd.edu/rgroups/ripe/perg/abp/mech.htm>; (s) E. Kashy, S.J. Gaff, N.H. Pawley, W.L. Stretch, S.L. Wolfe, D.J. Morrissey, Y. Tsai, "Conceptual questions in computer-assisted assignments," Am. J. Phys. **63**, 1000-1005(1995).

67. D. Griffiths, "Milliken Lecture 1997: Is there a text in this class?," *Am. J. Phys.* **65**, 1141-1143 (1997).

68. R. Karplus, "Science Teaching and the Development of Reasoning," J. Res. Sci. Teach. **14**, 169-175 (1977).

69. M. Csikszentmihalyi, "Literacy and Intrinsic Motivation," *Daedalus* **119**(2), 115-140 (1990): "The chief impediments to learning are not cognitive. It is not that students cannot learn; it is that they do not wish to. If educators invested a fraction of the energy they now spend trying to transmit information in trying to stimulate the students' enjoyment of learning, we could achieve much better results." Other references on motivation are T.W. Malone, "Toward a Theory of Intrinsically Motivating Instruction, Cog. Sci. **4**, 333-369 (1981); T. Malone and M. Lepper, "Making Learning Fun: A taxonomy of intrinsic motivations for learning," in *Aptitude, Learning, and Instruction:* Vol 3. *Conative and Affective Process Analysis*" (Erlbaum, 1987); M. Lepper, "Microcomputers in Education: Motivational and Social Issues," Am. Psychologist **40**, 1-18 (1985).

70. S. Tobias, *They're Not Dumb, They're Different: Stalking the Second Tier* (Research Corporation, Tucson, AZ, 1990).

71. E. Cullota, "The Calculus of Educational Reform," Scienc*e* **255**, 1060-1062 (1992).

72. (a) F. Reif and J.H. Larkin, "Cognition in Scientific and Everyday Domains: Comparison and Learning Implications," J. Res. Sci. Teach. **28**, 733-760 (1991). (b) D. Hammer, "Epistemological beliefs in introductory physics," Cogn. Instruct. **12**, 151-183 (1994).

73. Instruments for the measurement of student beliefs about the nature of science and learning are under development : (a) I. Halloun, "Views About Science and Physics Achievement: The VASS Story," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE*, ed. by E.F. Redish and J.S. Rigden, (AIP, Woodbury, NY, 1997). p. 605 - 613; I. Halloun and D. Hestenes, "Interpreting VASS Dimensions and Profiles," Sci. and Ed., in press; (b) E.F. Redish, R.N. Steinberg, and J.M. Saul, "The Distribution and Change of Student Expectations in Introductory Physics," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE*, ed. by E.F. Redish and J.S. Rigden, (AIP, Woodbury, NY, 1997). p. 689-697; E.F. Redish, J.M. Saul, and R.N. Steinberg, "Student expectations in introductory physics," Am. J. Phys. **66**, 212-224 (1998); (c) I. Novodvorsky and C. Johnson, "Computer Delivery of an Instrument to Assess Students' Attitudes Towards Science," AAPT Announcer **24**(4), 77 (1994).

74. H. Lucido, "Coaching Physics," Phys. Teach. **26**, 333-340 (1988); "Physics for Mastery," *ibid.*, **30**, 96-101 (1992).

75. K. A. Ericsson, R.T. Krampe, and C. Tesch-Romer, "The Role of Deliberate Practice in the Acquisition of Expert Performance," Psychological Rev. **100**(3), 363-406, 1993; K.A. Ericsson and N. Charness, "Expert performance, its structure and acquisition, Am. Psychologist **49**, 725-747 (1994).

76. R.R. Hake, "Towards Mastery of Newtonian Mechanics by the Average Student," AAPT Announcer **24**(1), 23 (1994).

77. For the IU94S (95S) courses of Table, the FCI pretest average for those 43 (45) students who had *not* taken high-school physics was 32% (37%). The FCI pretest average for the 123 (164) students who *had* taken high-school physics was 42% (43%). These translate into normalized gains for the high-school physics courses of g = 0.15 (0.09). It is conceivable that these low g's reflect a rapid dissipation of conceptual understanding in the years following the typical high-school course, as might be expected if only incoherent and loosely related bits of physics understanding (refs. 1b; 2a; 43f,g) are acquired. More research on knowledge changes over decade-long periods (see, e.g., ref. 73) would be of value.

78. For the IU95S course, we administered Indiana University's *Math Skills Assessment* (MSA) test (mostly simple algebra) to all 209 students (average level – midway between sophomore and junior). The MSA is normally given to incoming freshman at Indiana. The average score of our students was 64%, close to the 62% usually taken to indicate that a student is at risk in substantive introductory math and science courses. In addition, although about 70% of students entering the non-calculus-based IU courses

of Table Ic have completed a university calculus course, almost none seems to have the foggiest notion of the graphical meaning of a derivative or integral. Similar calculus illiteracy is commonly found among students in calculus-based introductory physics courses at IU. In my judgment, these calculus interpretations are essential to the crucial operational definitions (refs. 20-22,43d,50b,62) of instantaneous position, velocity, and acceleration: the term "substantive non-calculus-based mechanics course" is an oxymoron. For an attempt to get calculus ideas across with MBL see R.R. Hake, "A Microcomputer-Based SDI Lab Emphasizing the Graphical Interpretation of the Derivative and Integral," AAPT Announcer **28**(2), xx (1998).

79. A. B. Arons, "Uses of the Past: Reflections on United States Physics Curriculum Development, 1955 to 1990, Interchange **24**(1&2), 105-128 (1993): "....although substantial improvements were achieved in the development of hands-on primary and secondary curricula, their effectiveness was in large part negated by logistic problems and *deficiencies in teacher preparation*," (our *italics*); "Improvement of Physics Teaching in the Heyday of the 1960's," *Conference on the Introductory Physics Course on the occasion of the retirement of Robert Resnick,* Jack Wilson, ed. (Wiley, 1997) p. 13-20.

80. L. C. McDermott, "A perspective on teacher preparation in physics and other sciences: The need for special science courses for teachers," Am. J. Phys. **58**, 734-742(1990).

81. J.I. Goodlad, *Teachers For Our Nation's Schools* (Jossey-Bass, 1990) : "Few matters are more important than the quality of the teachers in our nation's schools. Few matters are as neglected.... A central thesis of this book is that there is a natural connection between good teachers and good schools and that this connection has been largely ignored....*It is folly to assume that schools can be exemplary when their stewards are ill-prepared*." (Our *italics*.)

82. (a) Case studies of seemingly successful approaches to organizational problems in colleges and universities have been reported by S. Tobias in *Revitalizing Undergraduate Science: Why Some Things Work and Most Don't* (Research Corporation, Tucson, AZ, 1992). (b) For the high-school counterpart see T. R. Sizer's *Horace's Compromise* [Houghton Mifflin (HM), 1985], *Horace's School* (HM, 1992), *Horace's Hope* (HM, 1996).

83. R.R. Hake, R. Wakeland, A. Bhattacharyya, and R. Sirochman, "Assessment of Individual Student Performance in an Introductory Mechanics Course," AAPT Announcer **24** (4), 76 (1994). Scatter plots of absolute gains vs pretest scores for all students in a class delineate relatively low-g and high-g students. We discuss various diagnostic tests (mechanics, mathematics, and spatial visualization) given to incoming students which might be used to recognize *potential* "low gainers" and thus initiate helpful intervention.

84. J. Epstein, (a) private communication, 12/97 (b) "Cognitive Development in an Integrated Mathematics and Science Program" J. Coll. Sci. Teach., Dec. 1997/Jan. 1998, p. 194-201; (c) "What Is the Real Level of Our Students, or What Do Diagnostic Tests Really Measure?" preprint, 1998; (d) O.G. Kolodiy and J. Epstein, *Integrated Mathematics & Science - A Discovery Approach* (Burgess Press International, 1982).

85. T. O. Pride, S. Vokos, L.C. McDermott, "The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems," Am. J. Phys. **66**, 147-157 (1998).

86. D. Meltzer, "Individual Student Variability in Conceptual Learning Gains: Search for Determining Factors," *AAPT Announcer* **27**(4), 89 (1998).

87. U. Treisman, "Studying Students Studying Calculus: A Look at the Lives of Minority Mathematics Students in College," Coll. Math. Jour. **23**(5), 362-372 (1992).

88. R. Ehrlich, private communications, 12/97, 4/98.

89. D. Hestenes, private communication, 8/97.