

# Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender, High-School Physics, and Pretest Scores on Mathematics and Spatial Visualization. \* †

Richard R. Hake

Indiana University (Emeritus), 24245 Hatteras Street, Woodland Hills, CA 91367

<rrhake@earthlink.net>

## I. INTRODUCTION.

In a previous survey (Hake 1998a,b; 2002a) the correlation of the average normalized gain  $\langle g \rangle$  on the FCI (Hestenes et al. 1992, Halloun et al. 1995) or Mechanics Diagnostic (Halloun & Hestenes 1985a,b) test for 62 courses with  $\% \langle \text{pretest} \rangle$  was a very low +0.02. However, open research questions remain as to the existence of "hidden variables," a term evidently first used in an educational sense by Meltzer (2001). Hidden variables that might influence FCI  $\langle g \rangle$ 's are, e.g., the averages over a class of gender, math proficiency, spatial visualization ability, completion of high-school physics courses, scientific reasoning skills, physics aptitude, personality type, motivation, socio-economic level, ethnicity, IQ, SAT, and GPA. One approach to this question is to investigate the relationship of *individual* student learning gains  $g$  with such variables for *single courses* (Hake et al. 1994, Meltzer 2001), as is done below. One of the goals of such work is to discover student-profile characteristics or diagnostic tests that might alert teachers to potential low- $g$  students. If such were known then corrective actions might be taken early in the course so as to improve its overall effectiveness.

---

\*Partially supported by NSF Grant DUE/MDR-9253965.

† The reference is R.R. Hake, "Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender, High-School Physics, and Pretest Scores on Mathematics and Spatial Visualization," submitted to the Physics Education Research Conference; Boise, Idaho; August 2002. Online at < <http://www.arxiv.org> > and also as ref. 22 at < <http://www.physics.indiana.edu/~hake> >.

© Richard R. Hake, 6/26/02. Permission to copy or disseminate all or part of this material is granted provided that the copies are not made or distributed for commercial advantage, and the copyright and its date appear. To disseminate otherwise, to republish, or to place at another website (instead of linking to the above URL) requires written permission.

## II. CHARACTERISTICS OF THE COURSE

In the present article I revisit the data for a course "P201" that I team-taught with Professor Rick VanKooten at Indiana University in the Spring of 1995. This course is an "algebra-based course" (an obvious oxymoron for any substantive mechanics course) primarily for premeds and health professionals. The traditional topics for P201 were "covered": Newtonian mechanics, work and energy, momentum and collisions, solids & fluids, vibrations, and sound. The course is designated "IU95S" in Table Ic & Iic of Hake (1998b). According to those tables, for this course:

(a)  $\langle g \rangle = 0.60$  as determined from the  $\langle \% \text{posttest} \rangle = 77 \pm 15\text{sd}$  and  $\langle \% \text{pretest} \rangle = 42 \pm 15\text{sd}$  for 209 students, each of whom took both the pretest and posttest. Here and hereafter in this paper the number after the " $\pm$ " will always be the standard deviation (sd);

(b) the original Force Concept Inventory (FCI) (Hestenes et al. 1992) was used, except that minor changes were made in the wording of seven of the questions so as to remove ambiguities. Neither the scores nor the point biserial coefficients for those questions showed significant changes from the IU93S results in which the test was the original FCI;

(c) the Kuder Richardson reliability coefficient KR-20 for the FCI posttest was 0.81;

(d) interactive engagement strategies included Socratic Dialogue Inducing Labs (Hake 1992), a few Microcomputer-Based Labs (Thornton & Solloff 1990), Concept Tests (Mazur 1997), and Minute Papers (Schwartz 1983);

(e) the rather standard introductory physics text by Serway & Faughn (*College Physics*, 4th edition) was utilized.

## III. SEARCH FOR "HIDDEN VARIABLES"

In the research discussed below we consider the scores of only the 203 students (93 females and 110 males) who took *all four diagnostic tests*: the FCI pretest, FCI posttest, a math test, and a spatial visualization test. Thus the group considered does not include all the 209 students considered in Hake (1998a,b). Both the math test and the spatial test were given to the students at the start of the course. The math test was the Indiana University "Math Skills Assessment" (MSA) test, mostly simple algebra. As of 1995, the MSA was normally given to incoming freshman at Indiana. The spatial test was the Purdue Spatial Visualization (Rotations) test (Guay 1977).

Table I summarizes the data of the present study. In the table **F** = Female, **M** = Male, and **A** = All students. Consistent with Hake (1998a,b; 2001a, 2002a) the following parameters are utilized:

(1) **g** is the *single-student* normalized gain, defined as:

$$g = \frac{\% \text{Gain}}{\% \text{Gain}_{\max}} \dots \dots \dots (1a)$$

$$g = (\% \text{posttest} - \% \text{pretest}) / (100 - \% \text{pretest}) \dots \dots \dots (1b)$$

(2) **<g>** is the course average normalized gain, defined as the actual average gain, %<Gain>, divided by the maximum possible actual average gain, %<Gain><sub>max</sub>:

$$\langle g \rangle = \frac{\% \langle \text{Gain} \rangle}{\% \langle \text{Gain} \rangle_{\max}} \dots \dots \dots (2a)$$

$$\langle g \rangle = (\% \langle \text{posttest} \rangle - \% \langle \text{pretest} \rangle) / (100 - \% \langle \text{pretest} \rangle) \dots \dots \dots (2b)$$

where %<posttest> and %<pretest> are the final (posttest) and initial (pretest) class percentage averages.

(3) **g-ave** is the course average normalized gain calculated as the average of the single-student normalized gains  $g_i$ :

$$g\text{-ave} = (1/N) \sum_i g_i = (1/N) \sum_i [ (\% \text{post}_i - \% \text{pre}_i) / (100\% - \% \text{pre}_i) ] \dots \dots \dots (3)$$

where N is the number of students taking both the pre- and post-tests and the summation is over all N students. The advantage of using <g> rather than g-ave in a survey of course performance (Hake 1998a,b) is discussed in Hake (2001a). In the present work g-ave has an advantage in allowing calculation of standard deviations (sd's) of g for students within a single class.

(4) "**d**" is Cohen's (1988) "effect size" defined as

$$d = (m_A - m_B) / [ (sd_A^2 + sd_B^2) / 2 ]^{1/2} \dots \dots \dots (4)$$

where  $m_A$  and  $m_B$  are population means expressed in the raw (original measurement) unit, and where the denominator is the root mean square of standard deviations for the A- and B-group means, sometimes called the "pooled standard deviation." Cohen's (1988, p. 24) very rough rule of thumb – based on typical results in social science research – is that  $d = 0.2, 0.5, 0.8$  imply respectively "small," "medium," and "large" effects, but Cohen cautions that the adjectives "are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation."

(5) **HSP** stands for High School Physics and  $\langle g_h \rangle$  is a *hypothetical*  $\langle g \rangle$  for high-school courses attended by the IU95S students as calculated from

$$\langle g_h \rangle = ( \% \langle \text{pre-HSP} \rangle - \% \langle \text{preNoHSP} \rangle ) / ( 100 - \% \langle \text{preNoHSP} \rangle ) \dots \dots \dots (5)$$

as discussed in (Hake 2000). Here I assume, as a rough approximation, that the HSP graduates would have averaged  $\% \langle \text{Pre-HSP} \rangle$  just after HSP, and would have averaged  $\% \langle \text{preNoHSP} \rangle$ , just before they took HSP.

Table I. Parameters for Course IU95S.

	A	E	M	d(M-F)	HSP	No HSP	d(HSP)
$\langle g \rangle$	0.604	0.540	0.669	—	—	—	—
$g\text{-ave}$	0.622±0.23	0.543±0.21	0.688±0.22	0.68	0.632±0.23	0.590±.21	0.19
$\% \text{math-ave}$	64.5±18	61.5±19	67.1±16	0.31	65.9±18	60.0±17	0.340
$\% \text{spatial-ave}$	64.5±19	56.8±17	70.9±18	0.82	65.0±18	62.7±20	0.120
$\langle \% \text{post} \rangle$	76.9±15	70.5±15	82.3±14	0.84	77.8±15	73.8±14	0.16
$\langle \% \text{pre} \rangle$	41.6±15	35.9±13	46.5±15	0.76	43.3±15	36.2±12	0.52
$\langle \% \text{gain} \rangle$	35.2±14	34.6±14	35.8±15	0.08	34.5±13	37.6±15	-0.22
$g\text{-ave (HSP)}$	0.632±0.23	0.543±0.21	0.709±0.22	0.77	—	—	—
$g\text{-ave (No HSP)}$	0.590±0.14	0.544±0.21	0.625±0.21	0.39	—	—	—
$g_h$	0.11	—	—	—	—	—	—

Among the interesting features of Table I are, in order of descending rows:

a. The course average normalized gain  $\langle g \rangle = 0.604$  as calculated from the averages of the pre- and post-tests (Eq. 2a,b) is 3% lower than the course average normalized gain  $g\text{-ave} = 0.622$  calculated from the average of the single-student normalized gains (Eq. 3). This difference is consistent with Hake (1998a, footnote 46), where it is shown that  $[ g\text{-ave} - \langle g \rangle ]$  is proportional to the  $g_j$ -weighted average of the deviations  $(\text{pre}_j - \langle \text{pre} \rangle)$ . Since the average of  $(\text{pre}_j - \langle \text{pre} \rangle)$  is zero, a low  $[ g\text{-ave} - \langle g \rangle ]$  implies a low correlation between  $g_j$  and  $\text{pre}_j$  for individual students, just as there is a low correlation between  $\langle g \rangle$  and  $\langle \text{pre} \rangle$  for courses as indicated in the

"Introduction." The fact that these two types of averages are generally within 5% of one another for classes of about 20 students or more is a consequence of the relatively low correlation of single-student  $g$ 's with pretests scores. Table II indicates that for IU95S, that correlation is a relatively low +0.32.

b. The average normalized gain  $\langle g \rangle = 0.604$  is somewhat larger than the average  $\langle \langle g \rangle \rangle_{48IE} = 0.48 \pm 0.14$  (std dev) of forty-eight Interactive Engagement courses ( $N = 4458$ ) surveyed by Hake (1998a,b). The second row of the table indicates that students who had completed High School Physics (HSP) achieved a larger  $g$ -ave than students who did not, but the effect size is only 0.19 and the very low hypothesized normalized gain  $g_h = 0.11$  (bottom row) [less than the  $\langle g \rangle_{14T-ave} = 0.23 \pm 0.04$  of 14 Traditional courses ( $N = 2084$ )] suggest that the  $g$ -ave differential may be due more to the fact that high-school students who take physics tend to be more scientifically oriented than to the education per se received in their high-school physics courses. (As usual, correlation doesn't necessarily indicate causation.) The low  $g_h$  may reflect:

- (1) A rapid decrease in physics understanding in the years following HSP, as might be expected if only incoherent and loosely related bits of physics understanding had been acquired.
- (2) A failure of HSP to impart much understanding of physics in the first place.
- (3) Some combination of "1" and "2".

In any case, the results suggest the ineffectiveness of HSP to promote long-term conceptual understanding and the need for improved physics education of K-12 teachers (Hake 2000; 2002b,c).

c. The average math-test score,  $\%math-ave$ , for All the students (average level – midway between sophomore and junior) was 64.5%, close to the 62% usually taken to indicate that a student is at risk in substantive introductory math and science courses. The average math scores for males were higher than those for females with a modest effect size of 0.31. Students who had completed High School Physics (HSP) achieved a larger  $\%math-ave$  than students who did not ( $d = 0.34$ ), but here again the differential is probably due more to the fact that high-school students who take physics tend to be more mathematically oriented than to the math education per se received in HSP.

d. Average scores on the spatial test,  $\%spatial-ave$ , for males were higher than those for females with a "large" effect size of 0.82, consistent with previous investigations (Pallrand & Seeber 1984, Linn & Peterson 1985, Lord 1987, Howe & Doody 1989, Baenenenger & Newcombe 1989, Halpern 1992, Friedman 1995,). This gender difference is often attributed to cultural factors (e.g., boys are more inclined than girls to play with construction toys, engage in motion-intensive sports, and play computer games). Consistent with this assumption, Baartmans & Sorby (1996) showed that women engineering students at Michigan Technological University could perform as well as men on spatial visualization tests if brought up to speed by a one-quarter (6 hr/week) visualization course based on the Baartmans & Sorby text.

e. Males achieved larger  $g-ave = 0.688$  than females with  $g-ave = 0.543$ , with a "close to large" effect size 0.68. As discussed in Hake (2002a); Henderson et al. (1999), McCullough (2000), Galileo Project (2002), and Meltzer (2001) have also reported  $g-ave_{males} > g-ave_{females}$  for some classes. For an abridged version of the Conceptual Survey of Electricity (Maloney et al. 2001), Meltzer calculated gender-difference effect sizes of 0.44 and 0.59 for two classes [ N = 59, 78 ] at Iowa State University, but observed no significant gender difference in two other classes [ N = 45, 37 ] at Southeastern Louisiana University.)

–

The scatter plots of Fig. 1, display the gender-related spread in the pre/post FCI data.

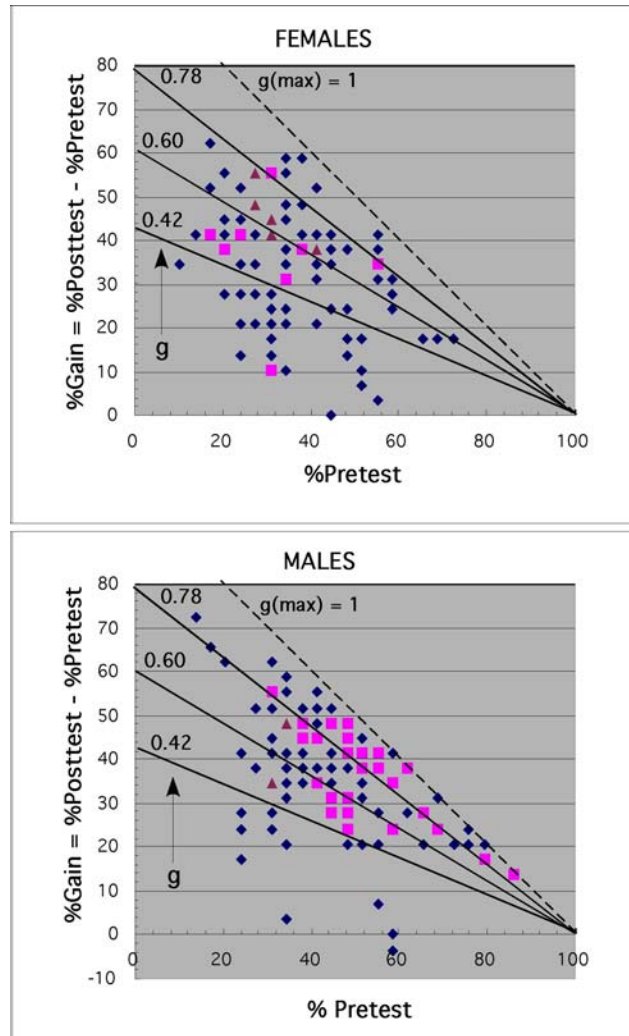


Fig. 1. Scatter plots of FCI %Gain vs. FCI pretest for the 93 females and 110 males of IU95S. The blue diamonds, purple squares, and brown triangles indicate, respectively, one, two, and three juxtaposed data points. The lines with negative slopes  $g = 0.42$  and  $g = 0.78$  form the boundaries of the high gain (HG) and low gain (LG) regions as explained in the text.

As in earlier work (Hake et al. 1994), I arbitrarily define for the present course the regions:

(a) "high gain" (HG) as  $g > 1.3 <g> = 1.3 (0.60) = 0.78$ ; containing 8 females and 39 males;

(b) "low gain" (LG) as  $g < 0.7 <g> = 0.7 (0.60) = 0.42$ ; containing 25 females and 12 males.

Thus, although females comprise 46% of the enrollment, they constitute 68% of the LG's and only 17 % of the HG's. For comparison, in an earlier course IU94S, Hake et al. (1994) found that

females, who comprised 47 % of the enrollment, constituted 71% of the LG's and (in sharp contrast to IU95S) 54% of the HG's. It's possible that the superior PER-based text by Reif (1994) and the slower pace (only mechanics was "covered") contributed to the higher percentage of female HG's in IU94S with its  $\langle g \rangle = 0.65$ .

#### IV. SEARCH FOR "CORRELATIONS"

According to educational psychologist Greg Schraw's (1998) PERC 1998 paper:

"There's a lot of snobbery, at least in quantitative research, and true . . . (physicists would require quotes around the education-specialist's use of "true") . . . experiments are always viewed as sort of the ideal. *Correlational studies are viewed as trash*, and quasi-experiments . . . (neither random selection nor assignment of subjects as in "true" experiments – see Cook & Campbell 1979). . . are somewhere in the middle. . . *there's too much correlational research out there already*, and we shouldn't promote any more of it than we have to." (*My italics.*)

So why am I about to talk "trash"? Because the correlations of Table II are part of an in-depth quasi-experimental study (Hake 1998a,b; 2002a) that, according to Schraw, "is exactly the kind of research that science educators need."

Table II. Correlations for IU95S.

Correlation Between	r		
	A	F	M
g & %math	0.36	0.21	0.44
g & %spatial	0.24	0.06	0.18
g & %pre	0.32	0.10	0.34
%post & %pre	0.58	0.48	0.55
%gain & %pre	-0.44	-0.41	-0.55

Table II shows correlations for All, Female, and Male students between single student's (a) normalized gain  $g$  and math, spatial, and pretest scores; (b) posttest and pretest scores; and (c) gain and pretest scores. Among features of interest are:



(a) The correlation  $r = 0.36$  between g's and math score is similar to the correlations observed by Meltzer (2001): 0.38 (N=45), 0.10 (N = 37 ), 0.46 (N = 59) , and 0.30 (N = 78) in four courses using an abridged version of the Conceptual Survey of Electricity (Maloney et al. 2001).

(b) The correlation  $r = 0.24$  between g's and spatial score would be classed by Cohen as "small." It is interesting that the correlation between female g's and spatial score is a very low 0.06.

(c) The correlations of +0.58 between post tests and pretest scores, and  $-0.44$  between gains and pretest score for *single students* are similar to those observed by Hake (1998a) for *courses*. The correlation of (%<posttest>) with (%<pretest>) was + 0.55, and the correlation of %<Gain> with %<pretest> was  $-0.49$ .

## V. CONCLUSIONS

The relatively modest correlations between g's and math, spatial, and FCI pretest scores (Table II) indicate that the tests used in this study cannot, by themselves, definitively identify potential low gainers. Likewise, the low effect size for dependence of g's on completion of high-school physics courses (Table I, second row) indicates that the completion of HSP is also, by itself, a poor indicator of university FCI gain. Nevertheless I suspect that relatively low scores on *all three* tests, or pathologically low scores on one or more of the tests might serve to identify potential LG's.

Possible intervention strategies for potential LG's are:

(a) Urge students with low diagnostic math scores to brush up. Some may need tutoring.

(b) Urge students who did not take HSP to exert extra effort, e.g., attend help sessions, join study groups, seek help from classmates and instructors.

(c) Interview students with relatively low scores on *all three* tests, or pathologically low scores on one or more of the tests to uncover serious cognitive or affective problems. If possible recommend a help menu.

(d) To enhance the spatial visualization of all students (not just potential LG's) meld some Baartmans & Sorby strategies into the course. Do more lab experiments such as the conical pendulum of Socratic Dialogue Inducing Lab # 3 "Circular Motion and Frictional Forces," (Hake 1998c) that require 2D representations of 3D motion.

A salient result of the present research is the demonstration of gender disparity in normalized gains [Table I (row 2) and Fig. 1]. The effect size  $d = 0.68$  is not far from the  $d = 0.8$  that Cohen loosely designates as "large." However, this gender effect size is eclipsed by the *very* large  $d = 2.43$  (Hake 2002a) for interactive engagement vs traditional courses in the survey of Hake (1998a,b). [Seven reasons for this unusually large  $d$  are given in Hake (2002a).] Thus, in my opinion, *effort to increase the degree of effective interactive engagement for ALL students should probably take a higher priority than effort to reduce gender disparity in FCI <g> values*, even despite the deplorable gender inequity in physics participation (Mallow & Hake 2002).

## References

- Baartmans, B. G. & S. A. Sorby. 1996. *Introduction to 3-D Spatial Visualization* (and accompanying Teacher's Resource Manual). Prentice Hall.
- Baenenenger, M., & N. Newcombe. 1989. "The role of experience in spatial test performance: A meta-analysis," *Sex Roles* 20: 327 -343.
- Cohen, J . 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum, 2nd ed.
- Cook, T. D. & D. T. Campbell. 1979. *Quasi-experimentation: Design & Analysis Issues for Field Settings*. Houghton Mifflin.
- Friedman, L. 1995. "The space factor in mathematics: gender differences," *Review of Educational Research* 65(1): 22-50.
- Galileo Project. 2002. A leading resource for teaching materials on the Web; online at < <http://galileo.harvard.edu/> >. Harvard FCI pre/post test results are at < <http://galileo.harvard.edu/galileo/lgm/pi/testdata.html> >.
- Guay, R.B. 1977 *Purdue Spatial Visualization Test: Rotations*. Purdue Research Foundation.
- Hake, R.R., R. Wakeland, A. Bhattacharyya, and R. Sirochman. 1994. "Assessment of individual student performance in an introductory mechanics course," *AAPT Announcer* 24 (4): 76.
- Hake, R.R.. 1992. "Socratic pedagogy in the introductory physics lab," *Phys. Teach.* 30 546-552.
- An updated version (4/27 /98) is online as ref. 3 at <<http://physics.indiana.edu/~sdi/>>. Socratic Dialogue Inducing Labs are online at the same URL.
- Hake, R.R. 1998a. "Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses." *Am. J. Phys.* 66(1): 64-74; also at < <http://www.physics.indiana.edu/~sdi/> >.

Hake, R.R. 1998b. "Interactive-engagement methods in introductory mechanics courses," submitted to *Physics Ed. Res. Supplement to Am. J. Phys.*; also at < <http://www.physics.indiana.edu/~sdi/> >. [ A crucial companion paper to Hake (1998a) – PER has no archival journal! ]

Hake, R.R. 1998c. SDI Lab # 3 "Circular Motion and Frictional Forces," online at < <http://www.physics.indiana.edu/~sdi> >. A teacher's guide is online (password protected) at Project Galileo (2002).

Hake, R.R. 2000. "Is it finally time to implement curriculum S? " *AAPT Announcer* 30 (4): 103 (2000); online as ref. 13 at < <http://www.physics.indiana.edu/~hake/> >.

Hake, R.R. 2001a. "Suggestions for Administering and Reporting Pre/Post Diagnostic Tests", unpublished; online as ref. 14 at < <http://physics.indiana.edu/~hake/> >.

Hake, R.R. 2001b. "Schwartz invented minute papers." Professional & Organization Development Network in Higher Education Discussion List; online at: < <http://listserv.nd.edu/cgi-bin/wa?A2=ind0105&L=pod&P=R4417> >

Hake, R.R. 2002a. "Lessons from the physics education reform effort," *Conservation Ecology* 5(2): 28; online at <<http://www.consecol.org/vol5/iss2/art28>>.

Hake, R.R. 2002b. "Physics First: Precursor to Science/Math Literacy for All? " to appear in the Summer 2002 issue of the American Physical Society's *Forum on Education Newsletter* < <http://www.aps.org/units/fed/index.html> > / "Forum newsletters" where "/" means "click on."; also online as ref. 19 at < <http://physics.indiana.edu/~hake/> >. For a more complete illustrated version see Hake (2002c).

Hake, R.R. 2002c. "Physics First: Opening Battle in the War on Science/Math Illiteracy? " Submitted to the *American Journal of Physics* on 27 June 2002; online as ref. 20 at < <http://physics.indiana.edu/~hake/> >.

Halloun, I. & D. Hestenes. 1985a. "The initial knowledge state of college physics students," *Am. J. Phys.* 53 1043-1055; online at < <http://modeling.la.asu.edu/R&E/Research.html> > .

- Halloun, I. & D. Hestenes. 1985b. "Common sense concepts about motion," *Am. J. Phys.* **53** 1056-1065; online at <<http://modeling.la.asu.edu/R&E/Research.html>>.
- Halloun, I., R.R. Hake, E.P Mosca, D. Hestenes. 1995. *Force Concept Inventory* (Revised, 1995); online (password protected) at <<http://modeling.la.asu.edu/R&E/Research.html>>.
- Halpern, D. 1992. *Sex Differences in Cognitive Abilities* (2nd edition). Erlbaum.
- Henderson, C.R., K. Heller, & P. Heller. 1999. "Common concerns about the Force Concept Inventory," *AAPT Announcer* **29**(4): 99 (1999); online at <<http://www.physics.umn.edu/groups/phised/Talks/talks.html>>.
- Hestenes, D., M. Wells, & G. Swackhamer, 1992. "Force Concept Inventory," *Phys. Teach.* **30** 141-158.
- Howe, A.C. and W. Doody. 1989. "Spatial visualization and sex-related differences in science achievement," *Science Educ.* **73** 703-709.
- Linn M.C. & A.C. Petersen. 1985. "Emergence and characterization of gender differences in spatial ability: A meta-analysis, *Child Development* **56**: 1479-1498.
- Lord, T.R. 1987. "A Look at Spatial Abilities in Undergraduate Women Science Majors," *Journal of Research in Science Teaching* **24**(8): 757-767; and references therein.
- Mallow, J.V. & R.R. Hake. 2002. "Gender Issues in Physics/Science Education (GIPSE) – Some Annotated References"; online at <<http://www.luc.edu/depts/physics/mallow.html>> and <<http://www.physics.indiana.edu/~hake>>; about 300 references and 200 hot-linked URL's.
- Maloney, D.P., T.L. O'Kuma, C.J. Hieggelke, and A. Van Heuvelen. 2001. "Surveying students' conceptual knowledge of electricity and magnetism," *Phys. Educ. Res., Am. J. Phys. Suppl.* **69**: S12-S23.
- Mazur, E. 1997. *Peer instruction: a user's manual*. Prentice Hall; online at <<http://galileo.harvard.edu/>>.
- McCullough, L. E. 2000. "Gender in physics: The past predicting the future? *AAPT Announcer* **30**(2): 81; online at <<http://physics.uwstout.edu/staff/mccullough/physicseduc.htm>>.

Meltzer, D.E. 2001. "The relationship between mathematics preparation and conceptual learning gains in physics: a possible 'hidden variable' in diagnostic pretest scores." Submitted to *Physics Ed. Res. Supplement to Am. J. Phys.*; online as ref. 5 at

< <http://www.physics.iastate.edu/per/articles/index.html> >.

Pallrand, G.J . & F. Seeber, 1984. "Spatial Ability and Achievement in Introductory Physics." *Journal of Research in College Teaching* 21(5): 507 -516.

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

Schraw, G. "Aspects of Validity in Quantitative Research," Physics Education Research Conference; Lincoln, Nebraska; online at < <http://physics.unl.edu/~rpeg/perc98/index.html> >.

Schwartz, C. 1983. *Minute Papers* as described in *ABCs of teaching with excellence: A Berkeley compendium of suggestions for teaching with excellence*, B.G. Davis, L. Wood, R.C. Wilson, University of California, Berkeley; online at

< <http://teaching.berkeley.edu/teaching.html>>. Minute Paper description online at < <http://teaching.berkeley.edu/compendium/suggestions/file95.html> >. For the origin of Minute Papers see Hake (2001b).

Thornton, R.K. & D. R. Sokoloff. 1990. "Learning motion concepts using real-time microcomputer-based laboratory tools," *Am. J. Phys.* 58(9): 858-867 .