

Using Matched Samples to Look for Sex Differences

Jennifer Blue, Miami University

bluejm@muohio.edu

Patricia Heller, University of Minnesota

helle002@tc.umn.edu

The reasons for observed differences in physics performance between men and women have yet to be clearly determined. This study asks the question: if men and women have a similar background at the start of an introductory physics course, will there be differences in how much physics they learn by the end of the course? To answer the question, a matched sample of men and women was studied. Statistical analysis of the post-tests reveals no significant differences between the men and women in the matched sample.

Introduction

There is considerable debate surrounding the causes of the observed differences in the physics performance of men and women. Some think that the differences are so often observed that they must be biological, and some think that differences in achievement must arise from social and cultural influences [1].

To add to this debate, this study asks the question: If there are minimal differences between men and women in their relevant physics background and initial performance when they start an introductory physics course designed to appeal to a broad population, will there be differences in how much physics they learn by the end of the course?

To answer this question, this study used matched samples. An attempt was made to control variables, to keep background and pretest performance constant and to have the only difference between men and women in matched pairs be their sex. If there are differences in the post-test performance of the men and women in the matched pairs, that will suggest that the differences are caused by their sex. Conversely, if there are no significant differences in the post-test performance of the matched pairs, that will suggest that sex of the students does not cause difference in performance.

The data in this study are drawn from a section of Physics 1251 offered at the University of Minnesota in the fall of 1993. The course is the first of a three-quarter sequence of calculus-based introductory physics, covering kinematics and dynamics. The students go to three lectures a week, plus one problem-solving session and one laboratory session. The problem solving sessions of this section incorporate cooperative group learning and an explicit problem solving strategy, both of which might help make it easier for women as well as men to experience some measure of success or enjoyment during their first physics course.

Methods

The students in the course filled out a demographic questionnaire, telling the following information: their year in college, whether they had taken physics in high school, the last high school math class they had taken, and their high school GPA. They also filled out a questionnaire about their locus of control over their own grades. Lastly, they took three ungraded pretests: the Force Concept Inventory (FCI) [2], a free response conceptual pretest, and a problem-solving pretest.

The FCI is a multiple-choice test. Each of the questions is worth one point. Questions are scored as correct if the

Newtonian response is selected, and incorrect if one of the non-Newtonian alternatives is selected.

The free response test consists of three questions: one about the acceleration of a ball rolling up and down the ramp written by the physics education group at the University of Washington [3], one about a car colliding with a bug as an example of Newton's Third Law, and one about the nature of forces and which forces make a car and passenger accelerate (designed by Patricia Heller, of the University of Minnesota physics education research and development group). For analysis purposes, these three questions are treated as six: (ramp, car and bug, nature of forces on the passenger, nature of forces on the car, why the passenger accelerates, and why the car accelerates. Well explained correct responses are given two points, partially correct responses are given one point, and incorrect responses are given no points. The maximum score for this test is therefore twelve points.

The problem-solving pretest is a mathematical reasoning problem, which assumes no physics knowledge. Student responses to that pretest are categorized as high, medium, and low.

A matched sample is chosen so that the performance of a group of men and women with minimal differences at the start of the course can be compared at the end of the course. This sample is necessarily small, for two main reasons. First, only about 22% of the students in the course are women. Second, much of the data are taken during lecture, and sometimes only half of the students in a given section attend on these days. Therefore, there are only complete data on 20 of the women.

The matched sample is made up of all the women on whom there are complete data and an equal number of men who are chosen to match them. The matched sample is chosen to eliminate as many

relevant, measurable differences as possible. The pairs are matched on eight measures: three pretest scores, three high school background characteristics, their year in college, and their locus of control over their grades. With so many variables, it is impossible to match all men to women whose characteristics are identical, but no matched pair has significantly large differences.

The performance at the end of the course is measured in three ways: the same FCI, the same free response test, and problems from the final exam. The FCI scores of men and women in the matched sample are compared with a matched-sample t-test. The free response test is analyzed as it had been when it was given as a pretest, and the total scores are analyzed using the Wilcoxon Matched-Pairs Signed-Ranks Test [4] to determine whether there are differences in the post-test conceptual ability of men and women in the matched sample.

The written problem solutions from the final exam are scored with a coding scheme developed at the University of Minnesota, based on research into how experts and novices solve problems. Solutions are coded on four measures of problem solving: their general approach, specific application of physics, logical progression, and use of appropriate mathematics. After each response is categorized, the categories are assigned numerical values between 0 and 10, where the category representing correct responses was always worth 10. The scores for the four measures are added for each problem, giving each student four scores out of 40 for a total possible score of 160. These total rank scores are analyzed using the Wilcoxon Matched-Pairs Signed-Ranks Test to determine whether there are differences in the post-test problem solving ability of men and women in the matched sample.

Table 1: Post-Test Results of the Matched Sample

Pair	FCI Post-Test			Free Response Post-Test			Problem-Solving Post-Test		
	Man	Woman	Δ	Man	Woman	Δ	Man	Woman	Δ
1	21	16	5	4	12	-8	120	124.29	-4.29
2	13	20	-7	0	5	-5	61.67	75.23	-13.56
3	17	14	3	4	7	-3	106.9	87.62	19.28
4	24	15	9	4	6	-2	145.71	70	75.71
5	14	24	-10	7	9	-2	145.46	131.68	13.78
6	22	22	0	2	3	-1	109.58	114.76	-5.18
7	25	20	5	1	2	-1	133.32	143.34	-10.02
8	24	21	3	7	8	-1	132.14	45.24	86.9
9	19	24	-5	2	3	-1	114.05	102.87	11.18
10	19	25	-6	3	3	0	95.25	37.14	58.11
11	22	25	-3	2	2	0	104.76	100.01	4.75
12	27	25	2	6	5	1	151.66	151.66	0
13	21	16	5	5	3	2	113.57	105.95	7.62
14	17	15	2	6	3	3	32.63	75.23	-42.6
15	22	22	0	11	8	3	120.01	123.33	-3.32
16	22	18	4	6	2	4	103.34	119.52	-16.18
17	23	14	9	6	2	4	94.05	92.6	1.45
18	22	15	7	10	5	5	101.43	58.81	42.62
19	17	24	-7	10	4	6	120.96	148.8	-27.84
20	26	22	4	11	3	8	143.34	129.27	14.07
Median	22	20.5	2.5	5	3	0	113.57	102.87	1.45

Results

The post-test results of the matched sample are in Table 1.

A matched-sample t-test indicates that there is no overall significant difference between the FCI scores of males and females ($t(df = 19) = -0.80, p = 0.430$) in the matched sample. In addition, Wilcoxon Matched Pairs Signed-Ranks Test shows that there is no overall significant difference between the free response test scores of males and females in the matched sample ($W(n = 19) = 68, p > 0.10$).

The score on four final exam problems are analyzed using Wilcoxon Matched Pairs Signed-Ranks Test, which looks for differences between

members of matched pairs. There is no difference in overall problem solving ($W(n=19) = 67, p > 0.10$) between men and women in matched pairs.

There is no difference in the problem solving ability of the men and women in either matched sample, as measured by their performance on their final exam problems. In addition, there is no difference in conceptual understanding as evidenced by scores on the FCI, overall scores on free response test, and scores on individual free response questions.

When the relevant measurable differences are removed from a sample of men and women in an introductory physics course designed to appeal to a

broad population, there are no differences in how much physics they learn by the end of the course.

Conclusions

When men and women are matched on high school backgrounds and pretest scores, there is no difference in post-test physics performance.

If there were biological sex differences in physics ability between the men and women in the matched sample, there would have been differences in their post-test scores. The fact that there were none provides supporting evidence for the hypothesis that the usual difference in performance seen in most studies is caused by social and cultural **gender** differences rather than biological **sex** difference.

Acknowledgements

Many thanks to Hans Courant, Kenneth Heller, Roger Jones, and Wayne Stein of the University of Minnesota for allowing close study of their physics classes.

Additional thanks are due to Anne Jones of Miami University for her help in the editing and preparation of this paper.

References

- [1] Kahle, J.B. and Meece, J. (1994). Research on gender issues in the classroom. In D. Gabel (Ed.), *Handbook of Research on Science Teaching and Learning*. Washington, DC: National Science Teachers Association.
- [2] Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *Physics Teacher*, 30, 141-158.
- [3] Trowbridge, D.E. & McDermott, L.C. (1981). Investigation of student understanding of the concept of acceleration in one dimension. *American Journal of Physics*, 49(1), 242-253.
- [4] Howell, D.C. (1992). *Statistical Methods for Psychology*. Boston: PWS-Kent.