University of Massachusetts Medical School
eScholarship@UMMS

Office of Institutional Research, Evaluation, and Assessment Publications and Presentations

Office of Educational Affairs

11-1-2004

The Use of "Effect Size" in Augmenting the Results of Significance Testing: A Comparison of Pre/Post Data from a Geriatric Interclerkship

Urip Purwono
University of Massachusetts Medical School

Susan V. Barrett
University of Massachusetts Medical School, Susan.Barrett@umassmed.edu

Michele M. Carlin
University of Massachusetts Medical School, Michele.Carlin@umassmed.edu

Mary L. Zanetti
University of Massachusetts Medical School, Mary.Zanetti@Umassmed.edu

Sarah M. McGee
University of Massachusetts Medical School, Sarah.McGee@umassmed.edu

Follow this and additional works at: http://escholarship.umassmed.edu/res_eval

Part of the Educational Assessment, Evaluation, and Research Commons, and the Medicine and Health Sciences Commons

Purwono, Urip; Barrett, Susan V.; Carlin, Michele M.; Zanetti, Mary L.; and McGee, Sarah M., "The Use of "Effect Size" in Augmenting the Results of Significance Testing: A Comparison of Pre/Post Data from a Geriatric Interclerkship" (2004). *Office of Institutional Research, Evaluation, and Assessment Publications and Presentations.* Paper 14.
http://escholarship.umassmed.edu/res_eval/14

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Office of Institutional Research, Evaluation, and Assessment Publications and Presentations by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

The Use of "Effect Size" in Augmenting the Results of Significance Testing: A Comparison of Pre/Post Data from a Geriatric Interclerkship

Urip Purwono, Susan V. Barrett, Michele M. Carlin, Mary L. Zanetti, Sarah McGee

University of Massachusetts Medical School

BACKGROUND

Most research in medical education, when examining the impact of an intervention, report findings based primarily on significance testing despite the controversy of its appropriate use. Moreover, the p-value used to determine rejection or acceptance of the null hypothesis tells nothing about the magnitude of the significance. Using a pre/post assessment of a Geriatric Interclerkship as a case study, this study examines the utility of "effect size" measures in augmenting significance testing results.

Effect size (ES) is a name given to a family of indices that measure the magnitude of a treatment effect. Unlike significance tests, these indices are independent of sample size. The formula used to calculate ES in this study follows:

$$d = M_1 - M_2 / \sigma_{pooled}$$

$$\sigma_{pooled} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$$

Where M_1 and M_2 are the means of the pre- and post- groups, respectively, and the pooled standard deviation is the square root of the average of the squared standard deviations (Cohen, 1988).

METHOD

A pre/post evaluation consisting of knowledge and attitude items was constructed and administered to students participating in the Geriatric Interclerkship during academic year 2003-2004. While percent of student change on attitude and knowledge items from pre- to post- was measured, matched pre/post data was examined in terms of significance testing (paired t-test) and one measure of "effect size" (Cohen's d).

RESULTS

Tables 1 and 2 highlight the percent of student pre/post change on attitude and knowledge items. Reliability of the instruments was .63 (knowledge test) and .82 (attitude scale). Ninety and ninety-two matched responses were collected to analyze clusters of attitude and knowledge items, respectively. Table 3 indicates the pre-post difference was significant for both the knowledge and attitude domains (t_{91} = 17.60 and t_{89} = 5.82, respectively; p <. 01). Table 4 shows the effect size for the knowledge domain was very high (d = 1.84) indicating a substantial change in student pre to post knowledge, thus complementing the result of the significance testing. On the contrary, the effect size for the attitude domain was small (d = .34), indicating a minimal change in student attitude.

Table 1: Percent of Students Changing on Attitude Questions From PRE to POST

	Neutral/Positive to Negative	Positive to Neutral	Negative to Negative	Neutral/Negative to Neutral	Positive to Positive	Neutral/Negative to Positive
Item 1		3%		11%	73%	12%
Item 2r	2%	1%		18%	66%	13%
Item 3r	6%	4%	34%	28%	18%	10%
Item 4		6%	1%	8%	80%	6%
Item 5r	2%	6%	4%	18%	63%	7%
Item 6r	13%	4%	21%	18%	26%	18%
Item 7	6%	8%	7%	32%	33%	14%
Item 8r	8%	2%	28%	28%	22%	12%
Item 9	13%	4%	28%	30%	19%	6%
Item 10r	1%	6%		4%	78%	11%
Item 11r		2%	2%	11%	73%	11%
Item 12r		3%		11%	78%	8%
Item 13r		4%	3%	7%	71%	14%
Item 14	2%	1%	1%	6%	79%	11%

Negative = Disagree, Strongly Disagree; Positive = Agree, Strongly Agree.
Items with reverse coding (2,3,5,6,8,10-13) reflect reverse scale.

Table 2: Please rate your level of competence in the following content areas:

	Right-Wrong	Wrong-Wrong	Right-Right	Wrong-Right
Item 1	1%	5%	89%	4%
Item 2		9%	14%	77%
Item 3	4%	49%	5%	41%
Item 4	11%	25%	29%	35%
Item 5		48%	8%	45%
Item 6	1%	9%	32%	59%
Item 7	10%	18%	51%	21%
Item 8	3%	32%	25%	40%
Item 9		38%	15%	47%
Item 10	8%	25%	49%	18%
Item 11	5%	72%	5%	17%
Item 12	9%	14%	58%	20%
Item 13	10%	32%	18%	40%
Item 14	2%	11%	70%	17%
Item 15	2%	3%	86%	9%
Item 16			99%	1%
Item 17	1%	3%	79%	16%
Item 18	1%		89%	10%
Item 19			99%	1%
Item 20	10%	8%	60%	23%
Item 21	2%	2%	88%	8%
Item 22	1%	1%	92%	5%
Item 23			93%	7%
Item 24	10%	14%	58%	18%
Item 25	15%	40%	29%	15%

Table 3: Paired Samples Test

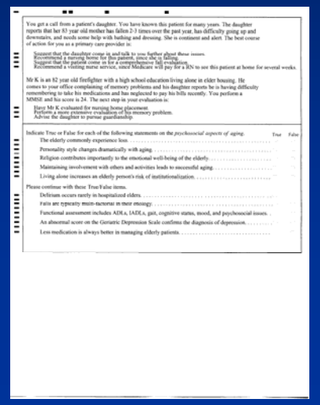
		Mean (SD)	N	Std. Error Mean	Paired Differences: Mean (SD)	Std. Error Mean	95% CI of the Difference		t	df	Sig. (2-tailed)
Pair 1	Post Knowledge	19.36 (2.72)	92	.284	4.88 (2.66)	.277	4.33	5.43	17.60	91	.000
	Pre Knowledge	14.48 (2.58)	92	.269							
Pair 2	Post Attitude	3.69 (.431)	90	.045	.16 (.25)	.027	.10	.21	5.82	89	.000
	Pre Attitude	3.54 (.453)	90	.048							

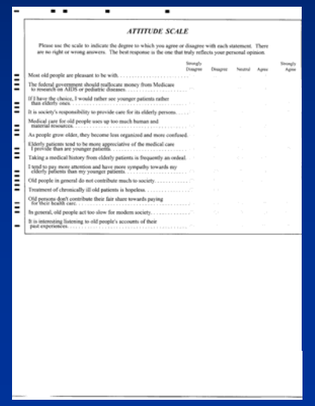
Table 4: Effect Size: Cohen's d

Domain	*d*	95% CI of Effect Size		Cohen's Standard
Knowledge	1.84	1.49	2.18	Large
Attitude	.34	0.04	0.63	Small

Evaluation Tools

Geriatrics Interclerkship
November 7, 2003 PRE/POST Evaluation





ATTITUDE SCALE

Geriatric Experience

CONCLUSION

The "effect size" provides additional practical information to the significance testing. The routine use of this analysis is recommended to enhance the quality of research and evaluation in medical education. More specifically, the results from this Geriatric Interclerkship could eventually be compared to the change in student knowledge and attitude in another Interclerkship experience. Therefore, the most important benefit of reporting effect sizes is that this information provides the researcher with a more standard tool that allows for meta-analysis across studies.