

FOSS Virtual Tutor: Comparison

Tim Weston 2_13_09

The comparison examined pre and post ASK assessments for students participating in “Wizard of Oz” (WOZ) FOSS tutoring for Magnetism and Electricity (M & E), and students who did not participate. Tutoring occurred 13 times from November 13, 2008 to January 22, 2009 with eight tutors. Each tutoring session was 20 minutes.

Sample

Two classrooms school participated, 46 students completed the pre assessment, 42 the post; 39 students received both assessments. Of these 39 students, 31 received WOZ tutoring, 8 were in the control condition in which they participated in regular FOSS lessons.

Measure

The assessment for M & E contained 12 items pre and 14 items post from the baseline/post version of the FOSS ASK assessment. (Two items were added to the post about electromagnets and will be retained in the next round of testing). Content covered the basic concepts found in the M & E lesson including magnets, flow of electricity, temporary magnets and magnetic polarity. All items (except two) call on students to give a written response or draw a picture; responses are graded using the scoring rubric provided by FOSS. Most items used a 0 – 3 scale, 2 items used a 0 -2 scale. Total possible for pre equaled 35 points; total possible for post equaled 40 points. We used the average of the two rater’s scores as the outcome measure.

Inter-rater reliability, internal consistency of measure

We assessed inter-rater reliability for total score for the two raters grading the pre and post assessments using an Intra-class Correlation Coefficient (ICC) which takes into account agreement and possible easiness/strictness (bias) between raters. For two raters the ICC is usually very close, or identical to a correlation coefficient. Two raters graded both the pre and post assessments for all students independently after practicing rating student responses with the rubric.

Both pre and post tests showed very high inter-rater reliability with an ICC = .96 for pre, and ICC = .94 for the post. Rater bias was negligible, with Rater 1 slightly stricter for the pre test (.16 on a 35 point scale) and Rater 2 slightly stricter on the post (.11 on a 40 point scale).

The internal consistency for both pre and post assessments was also very high with Cronbach’s α = .98 for pre, and α = .97 for post.

Analysis procedures

We compared students on total score for the ASK pre and post to learn if mean scores and pre/post gain differed according to: 1) if students participated in WOZ or not, 2) the number of WOZ sessions students completed with tutors. We also examined if outcomes varied by which tutor guided the WOZ program.

Statistical procedures included Repeated-measures ANOVA and Linear Regression.

WOZ Comparison

Tables 1 -2 show means for pre, post and gain for WOZ/Control, Number of Sessions

Table 1 Means for Pre, Post (Total Score) for *WOZ/Control* comparison

| | WOZ INTERVENTION | Mean | Std. Deviation | N |
|----------------------|------------------|-------|----------------|----|
| TOTAL PRE (AVERAGE) | Control | 13.63 | 5.80 | 8 |
| | WOZ | 14.77 | 3.90 | 31 |
| | Total | 14.54 | 4.29 | 39 |
| TOTAL POST (AVERAGE) | Control | 22.75 | 5.51 | 8 |
| | WOZ | 28.84 | 3.55 | 31 |
| | Total | 27.59 | 4.66 | 39 |
| GAIN | Control | 9.13 | 3.1 | 8 |
| | WOZ | 14.06 | 4.5 | 31 |

^ Analysis only includes students taking both tests

Table 2 Means for pre, post and gain for *Number of WOZ Sessions* variable

| | NUMBER OF SESSIONS | Mean | Std. Deviation | N |
|----------------------|--------------------|-------|----------------|----|
| TOTAL PRE (AVERAGE) | 0 | 13.63 | 5.80 | 8 |
| | 2 | 13.13 | 6.69 | 4 |
| | 3 | 14.14 | 3.52 | 7 |
| | 4 | 16.19 | 2.92 | 13 |
| | 5 | 14.33 | 4.02 | 6 |
| | 6 | 10.00 | . | 1 |
| | Total | 14.54 | 4.29 | 39 |
| TOTAL POST (AVERAGE) | 0 | 22.75 | 5.51 | 8 |
| | 2 | 26.25 | 2.22 | 4 |
| | 3 | 29.71 | 4.49 | 7 |
| | 4 | 30.08 | 1.85 | 13 |
| | 5 | 27.25 | 5.07 | 6 |
| | 6 | 26.50 | . | 1 |
| | Total | 27.59 | 4.66 | 39 |
| GAIN | 0 | 9.13 | 3.13 | 8 |
| | 2 | 13.13 | 6.69 | 4 |
| | 3 | 15.57 | 5.97 | 7 |
| | 4 | 13.88 | 2.98 | 13 |
| | 5 | 12.92 | 4.93 | 6 |
| | 6 | 16.50 | . | 1 |
| | Total | 13.05 | 4.68 | 39 |

Figures 1 -4 show means for total pre, post and gain for WOZ and Number of WOZ Sessions.

Figure 1 Mean Pre/Post Total Scores for WOZ/Control Comparison

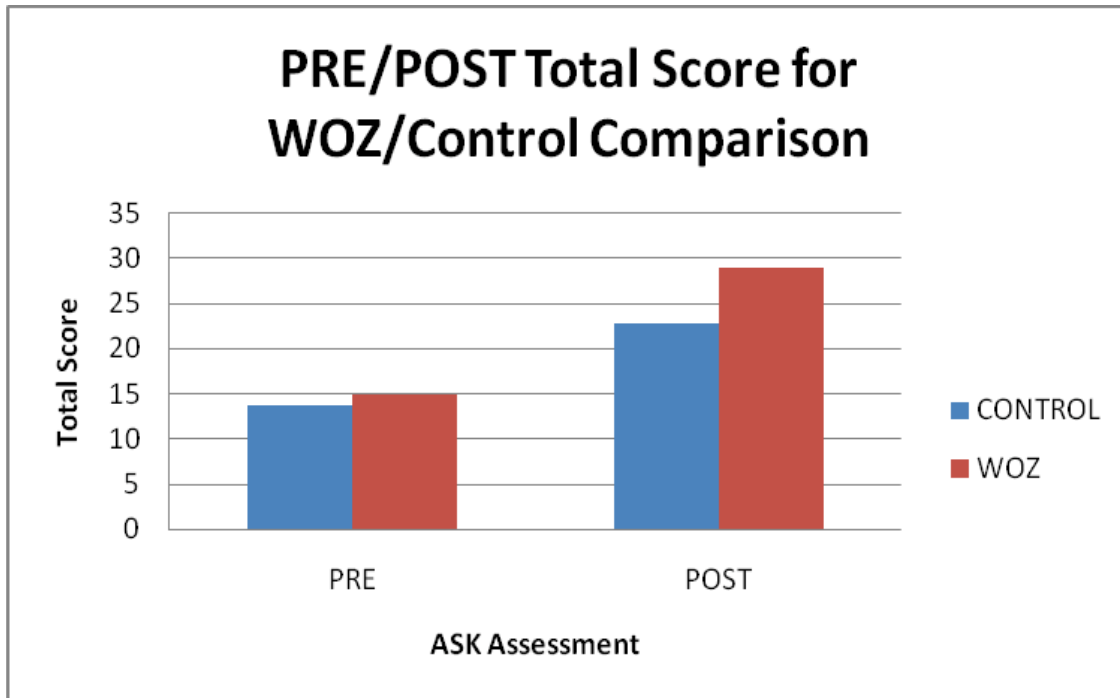


Figure 2 Mean Pre/Post Gain by group for WOZ/Control Comparison

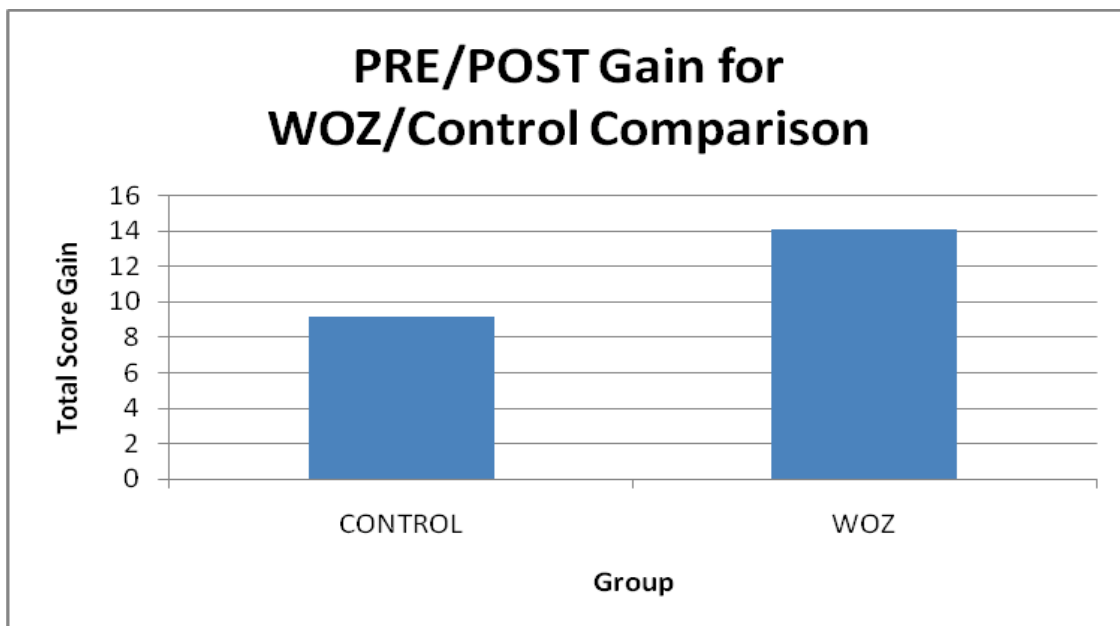


Figure 3 Mean pre/post gain for “Number of Sessions” variable

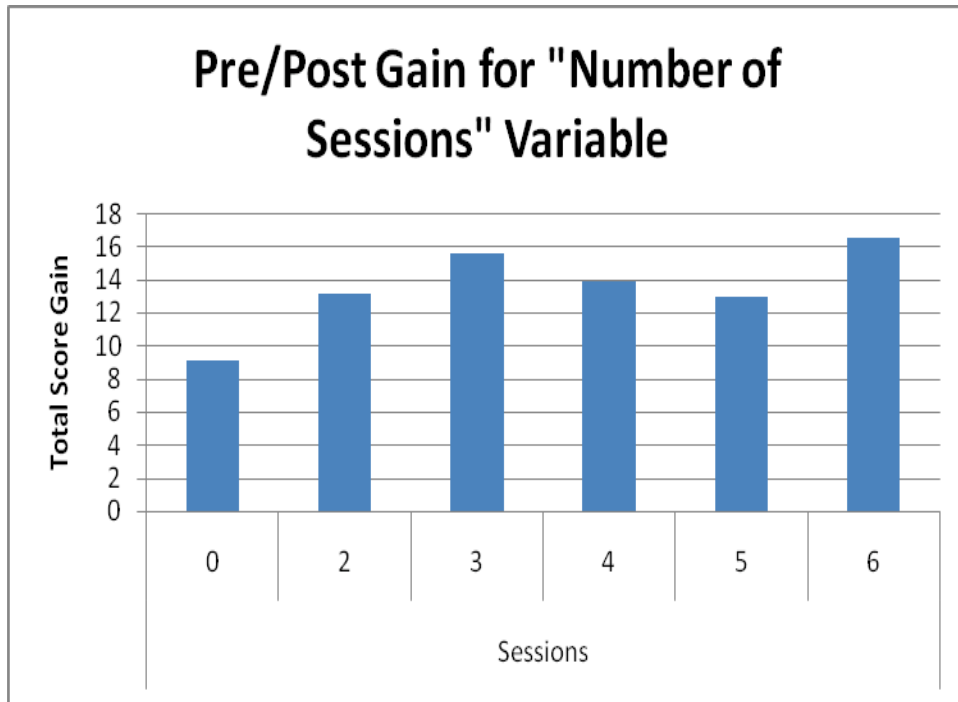
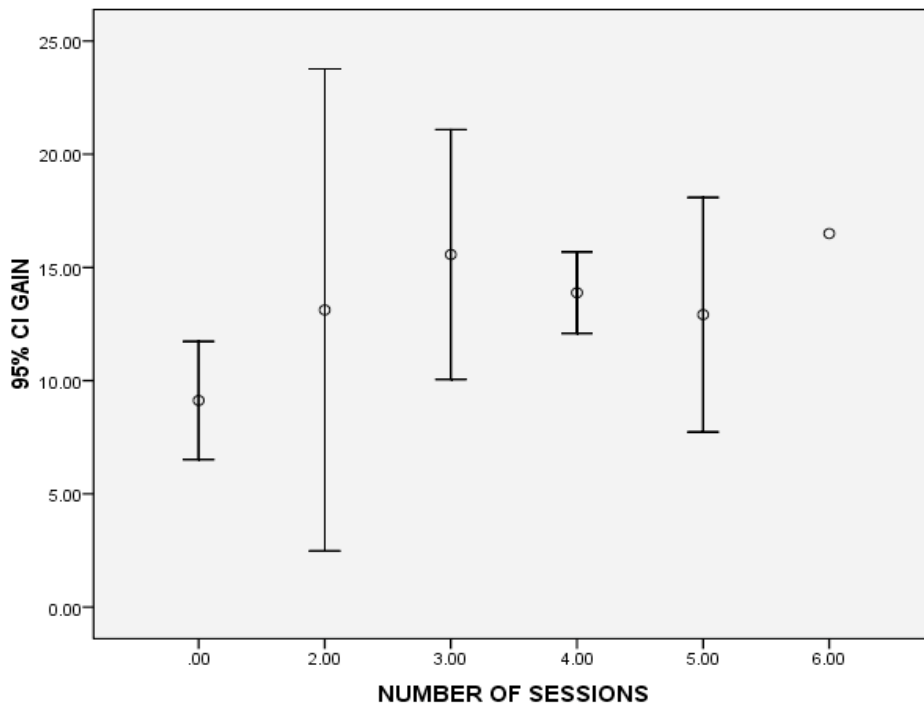


Figure 4 Dispersion for Pre/Post Gain for “Number of Sessions” variable



Statistical Analysis

We first tested the mean WOZ/Control difference using a Repeated-Measures ANOVA with Group (WOZ/Control) used as a fixed factor.

Tests of Within-Subjects Contrasts

| Source | PREPOST | Type III Sum of Squares | df | Mean Square | F | Sig. |
|----------------|---------|-------------------------|----|-------------|---------|------|
| PREPOST | Linear | 1709.781 | 1 | 1709.781 | 186.546 | .000 |
| PREPOST * WOZ | Linear | 77.576 | 1 | 77.576 | 8.464 | .006 |
| Error(PREPOST) | Linear | 339.123 | 37 | 9.165 | | |

The Pre/post by group interaction was significant at $F = 8.4$, $df 1, 34$, $p = .006^{**}$. This result indicates that students in the WOZ group gained significantly more than students in the control group.

We then used the “*Number of Sessions*” variable as a predictor in linear regression analysis with Pre test to learn if Number of Sessions predicted Post test with pre-score held constant.

Coefficients(a)

| Mode | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|------|--------------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 17.897 | 2.305 | | 7.763 | .000 |
| | Pre Test | .444 | .144 | .408 | 3.086 | .004 |
| | Number of Sessions | 1.081 | .344 | .415 | 3.139 | .003 |

a Dependent Variable: Post Test

$$R^2 = .378$$

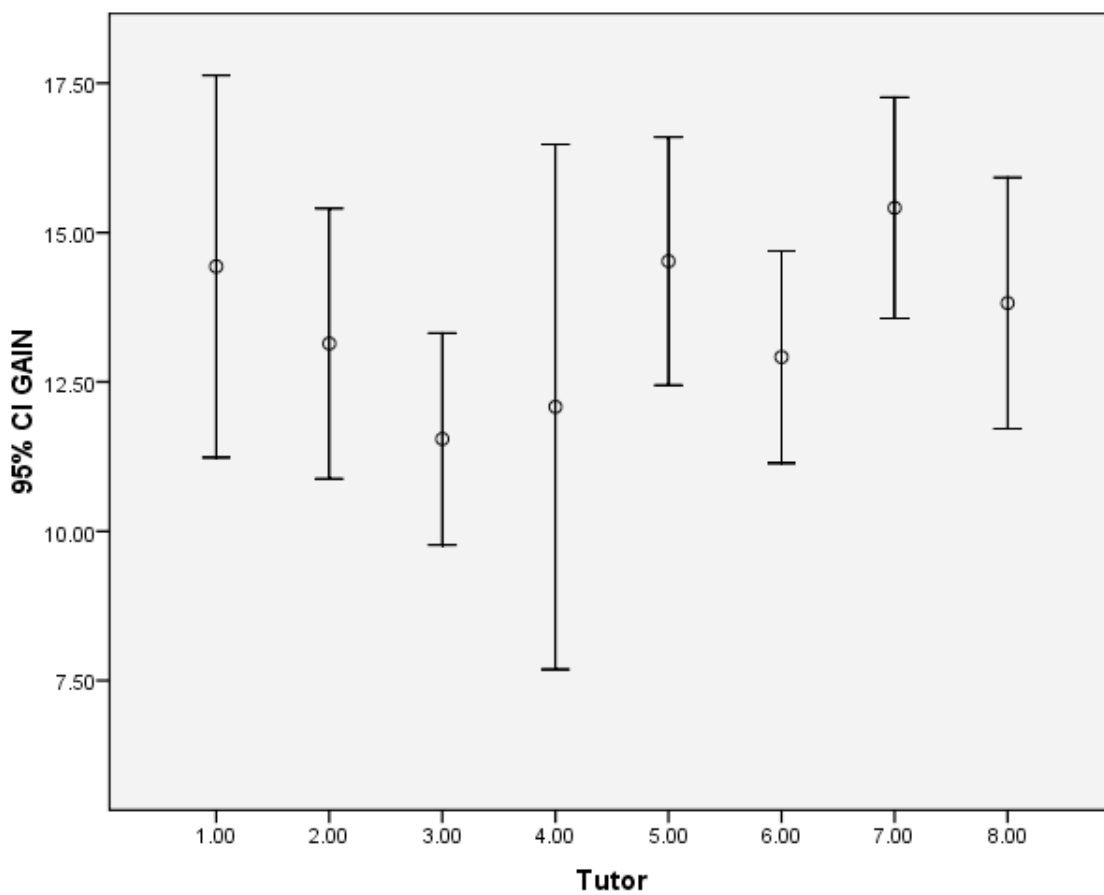
The number of sessions predicted *Post Test* with *Pre Test* held constant with $\beta = .415$, $t = 3.139$, $p = .003^{**}$. This result indicates that number of sessions had a significant linear relationship with post score.

Caveats: Looking at the figures 3 and 4 it should be noted that *Gain* dropped for the 4 and 5 session groups. Also, only one student received 6 sessions. When this student was removed from the analysis results did not change substantially with $t = 3.16$, $p = .003^{**}$.

Tutor and Class Variance

We examined the variability of outcomes due to tutors. Given the structure of the data (each student received tutoring from multiple tutors) it was difficult to compare and isolate tutor effects with the effects of other variables. However, we could examine the variability of pre/post gain associated with each tutor

Means and dispersion around means for each tutor are presented in figure 5.



Gain score varied with tutor accounted for approximately 7% of total variance in gain score.