DETERMINING THE EDUCATIONAL EFFECTIVENESS OF COMBINING VOICE-OVER PRESENTATIONS AND REVIEW GAMES: RESULTS

Prepared by: Marcos A. Sanchez-Gonzalez, MD, PhD

A. 3-Group Comparison (PPT vs. PPT+VOICE vs. edYOU)

1. *Statistical Analysis* - A one-way analysis of variance (ANOVA) was conducted to compare the effectiveness of the different educational strategies: slide decks only or traditional way (PPT), PPT plus voice-over (PPT+VOICE), and artificial intelligence platform (edYOU). If significant interactions were found, the corresponding post-hoc test using Fisher's Least Significant difference test was used as a follow-up. A priori, an alfa level of 0.05 was considered significant.
2. *Results* - Data are presented as mean ± s.e.m. Descriptive statistics are shown in Table 1. Students using PPT+VOICE and edYOU obtained higher quiz scores, difficult questions quiz scores, and percent of valid questions than PPT, albeit it was statistically significant ($P>0.05$). There was a statistically significant ($P < 0.001$) interaction in total such that students in the PPT+VOICE group spent less time in lectures (34.5 ± 13.5 hrs) and also the edYOU group (54.1 ± 13.7 hrs) compared to the PPT group. Despite the time difference between PPT+VOICE and the less time of edYOU of student spent in lectures (-21.6 ± 13.7 hrs.), it was not statistically significant ($P = 0.12$).

Table 1. Descriptive Statistics For the Different Educational Strategy Groups

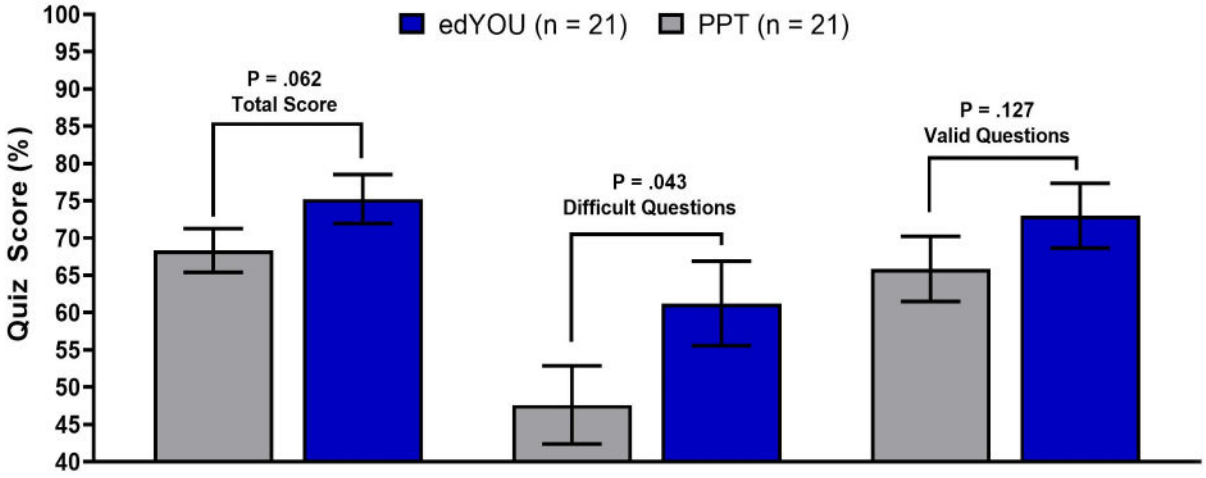
Group		Total Score (%)	*Difficult Questions Score (%)	*Valid Questions Score (%)	Total Time (Hours)
PPT	**N**	21	21	21	21
	Mean	68.33	47.62	65.87	74.76
	Minimum	45.00	0.00	33.33	0.73
	Maximum	90.00	85.71	91.67	163.99
	SEM	2.93	5.25	4.38	13.11
PPT + VOICE	**N**	20.00	20.00	20.00	20.00
	Mean	74.00	58.57	73.33	42.28
	Minimum	55.00	14.29	41.67	0.10
	Maximum	95.00	85.71	100.00	142.46
	SEM	2.70	5.38	3.99	8.33
edYOU	**N**	21.00	21.00	21.00	21.00
	Mean	75.24	61.22	73.02	20.64
	Minimum	45.00	14.29	33.33	0.13
	Maximum	95.00	100.00	100.00	105.32
	SEM	3.28	5.68	4.34	5.95
Total	**N**	62	62	62	62
	Mean	72.50	55.76	70.70	45.95
	Minimum	45.00	0.00	33.33	0.10
	Maximum	95.00	100.00	100.00	163.99
	SEM	1.74	3.18	2.45	6.18

*See item analysis section C

B. 2-Group Comparison (PPT vs. edYOU)

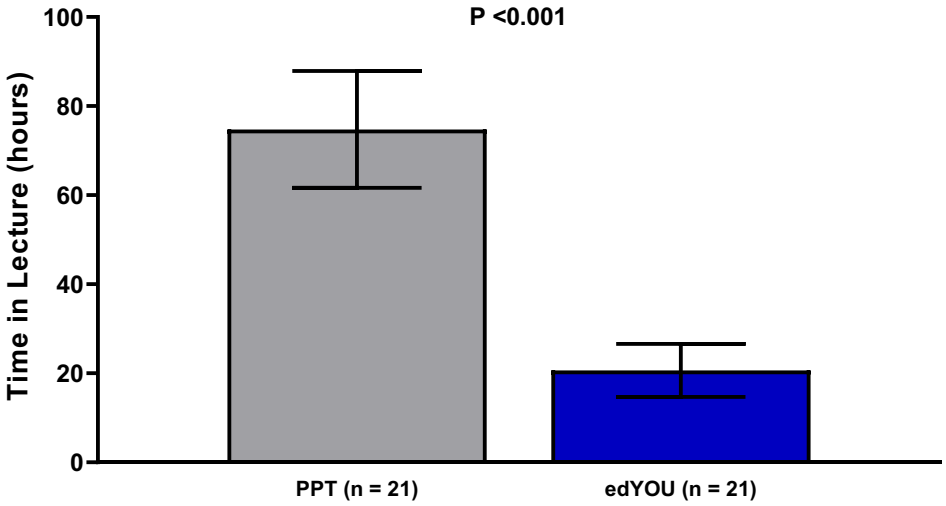
1. *Statistical Analysis* – Student's T-test was conducted to compare the effectiveness of the different educational strategies: slide decks only or traditional way (PPT), PPT plus versus artificial intelligence platform (edYOU). A priori, an alfa level of 0.05 was considered significant.
2. *Results* - Data are presented as mean ± s.e.m. Students using edYOU obtained higher quiz scores, difficult questions quiz scores, and percentage of valid questions than PPT (Figure 1). There was a statistically significant ($P < 0.001$) decrease in Time spent in lectures in the edYOU group (54.1 ± 13.7 hrs) compared to the PPT group (Figure 2).

Figure 1. Quiz Results in Response to Educational Strategy



Note: Slide decks only or traditional way, PPT; artificial intelligence platform, edYOU.

Figure 2. Comparison of the Time spent in the lectures



Note: Slide decks only or traditional way, PPT; artificial intelligence platform, edYOU.

C. *Item Analysis

The statistical strategy for conducting item analysis included *item difficulty*, *item discrimination*, and *point-biserial correlation R* (PBI) (Bibler Zaidi et al., 2018). Briefly, the *item difficulty* was calculated by looking into the proportion of the total learners answering correctly; this metric's target was 50-70% (valid difficult questions). The item discrimination accounts for the difference between the upper quartile (of total scores) who answered correctly vs. the lower quartile of those also answering the item correctly; the main goal was to avoid negative scores (invalid questions). Lastly, the PBI was performed via the correlation between test and item scores; the desired target was over 0.20. Items not meeting the criteria described above were considered non-valid and hence nullified.

Reference

Bibler Zaidi, N. L., Grob, K. L., Monrad, S. U., Holman, E. S., Gruppen, L. D., & Santen, S. A. (2018). Item Quality Improvement: What Determines a Good Question? Guidelines for Interpreting Item Analysis Reports. *Medical Science Educator*, *28*(1), 13-17. https://doi.org/10.1007/s40670-017-0506-1

END OF DOCUMENT