

Appendices

Appendix 0: A link to a shared storage space on a drive

The quantitative test data, item analysis, outputs for statistical models, all tables and figures as well as transcribed qualitative data, preliminary questionnaire analyses, etc. can be accessed at:

<https://drive.google.com/drive/folders/0Byuw22bnyFgQMF14bE1TUWtJT0E>

Appendix 1: Quantitative data analysis PILOT

Appendix 1A:

PILOT: Test scores comparison

Compare means (t test): PPT condition (M = 64.11, SD = 12.39) vs CBT condition (M = 62.94, SD = 13.49), $t_{(137)} = 1.06$, $p = 0.29$

Compare medians (Wilcoxon rank test, non-param): PPT condition (Median = 66, MAD = 14.83) vs CBT condition (Median = 64.18, MAD = 14.28), $V = 5443.5$, $p = 0.17$

The p values in both tests are above the standard alpha level ($\alpha < .05$).

Appendix 1B:

PILOT: Test scores by mode of administration and gender

Gender	Mode	N	Mean	SD	Median	Min	Max	Range
Female	Computer-based	88	63.8	13.9	64.8	25.2	90.5	65.4
Female	Pencil & paper	88	64.3	12.6	65.5	33.0	89.0	56.0
Male	Computer-based	50	61.3	12.7	63.2	36.2	85.1	48.9
Male	Pencil & paper	50	63.7	12.1	66.2	37.0	82.0	45.0

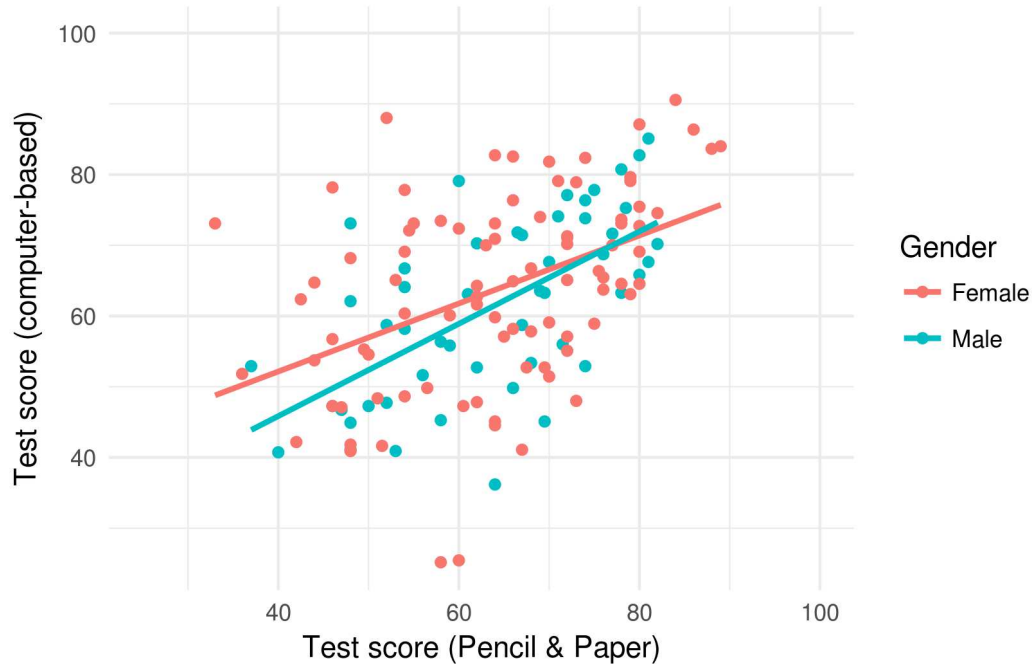
N = number of students, SD = standard deviation

Appendix 1C:

PILOT: Scatter plot

Test Scores in Two Modes of Administration

by mode, gender (pilot data)



Appendix 2: Quantitative Data Analysis Study 1

Appendix 2A:

Study 1 - Test scores by gender

Gender	Subtest	N	Mean	SD	Median	Min	Max	Range	Skew
Female	Destination	142	13.0	3.2	13.2	5.0	18	13.0	-0.5
Female	Vocabulary	142	14.8	3.2	15.2	1.0	20	19.0	-1.3
Male	Destination	86	12.9	2.8	13.2	3.5	18	14.5	-0.9
Male	Vocabulary	86	15.0	2.3	15.2	9.4	20	10.6	-0.3

N = number of students, SD = standard deviation

Appendix 2B:

Study 1 - Destination scores by group and mode

Group	Mode	N	Mean	SD	Median	Min	Max	Range	Skew
Computer first	Computer-based	58	11.7	2.8	11.9	5.0	16.7	11.7	-0.5
Pencil &	Computer-	56	14.1	2.7	14.6	6.5	18.0	11.5	-0.8

paper first	based								
Computer first	Pencil & paper	58	13.9	3.0	14.2	5.0	18.0	13.0	-0.9
Pencil & paper first	Pencil & paper	56	12.2	3.0	12.8	3.5	18.0	14.5	-0.7

Vocabulary scores by group and mode

Group	Mode	N	Mean	SD	Median	Min	Max	Range	Skew
Computer first	Computer-based	58	13.3	3.2	13.5	1	18.5	17.5	-1.6
Pencil & paper first	Computer-based	56	15.9	2.4	16.0	10	20.0	10.0	-0.4
Computer first	Pencil & paper	58	15.9	2.3	16.0	8	20.0	12.0	-1.4
Pencil & paper first	Pencil & paper	56	14.5	2.9	15.0	7	19.8	12.8	-0.4

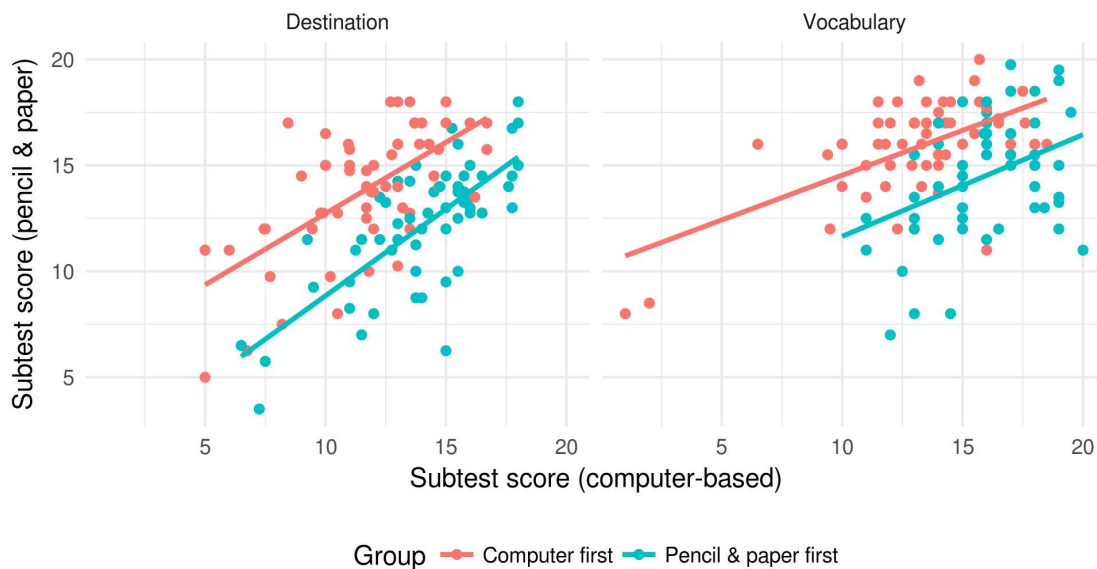
N = number of students, SD = standard deviation

Appendix 2C:

Study 1 - Scatter plot – Correlation of subtest scores in two modes of administration

Subtest Scores in Two Modes of Administration

by group (study 1)



Appendix 3: Quantitative Data Analysis Study 2

Appendix 3A:

Study 2 - Test scores by group

Group	Subtest	N	Mean	SD	Median	Min	Max	Range	Skew
Computer first	Destination	124	17.2	4.0	17	8	25.0	17.0	-0.1
Computer first	Vocabulary	124	15.8	4.4	16	4	24.5	20.5	-0.4
Pencil & paper first	Destination	128	15.9	5.1	16	2	25.0	23.0	-0.1
Pencil & paper first	Vocabulary	128	14.2	5.3	15	3	23.9	20.9	-0.3

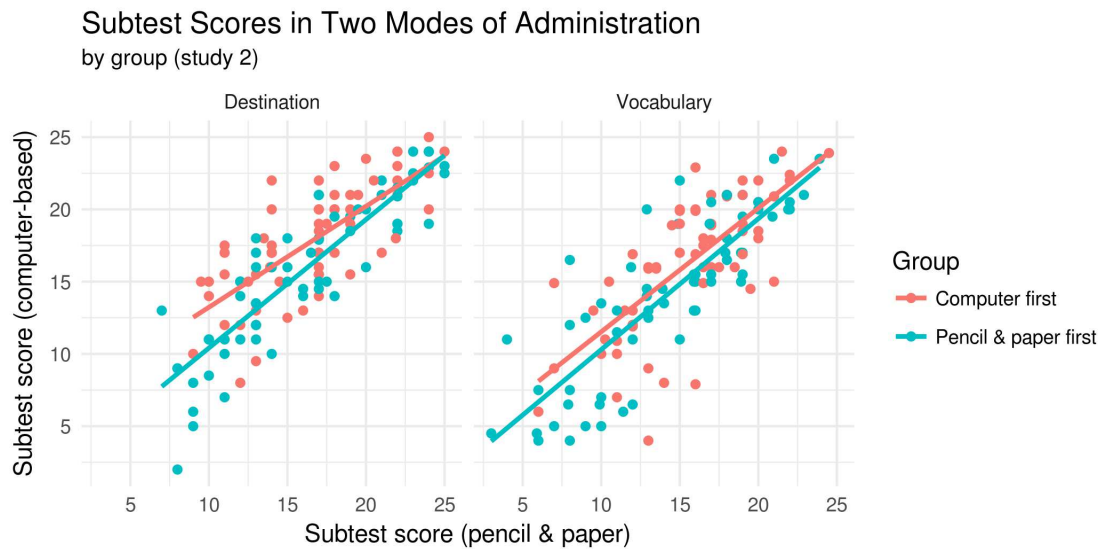
Appendix 3B:

Study 2 - Test scores by gender

Gender	Subtest	N	Mean	SD	Median	Min	Max	Range	Skew
Female	Destination	174	16.7	4.9	17.0	2	25.0	23.0	-0.3
Female	Vocabulary	174	15.0	4.9	15.9	3	24.5	21.5	-0.4
Male	Destination	78	16.2	4.1	17.0	5	25.0	20.0	-0.2
Male	Vocabulary	78	15.0	4.9	15.2	4	23.9	19.9	-0.4

Appendix 3C:

Study 2 - Scatter plot – Correlation of subtest scores in two modes of administration



Appendix 4: Qualitative Data Analysis

Appendix 4A

A list of themes and codes in alphabetical order:

Theme	Code	Theme	Code
Attitude	A	No Pen	NP
Changing answers	CA	No problems with handwriting	NHW
Cheating	Ch	No Results	NR
Classroom	Cl	No Technology	NT
Comfort	C	Noise	Noi
Computer fan	CF	Nothing	N
Concentration	Con	Orientation	O
Content-related	C-r	Performance-related	P-r
Difficulty	D	Physicality	Ph
Environment	E	Practical	P
Everything	Ev	Resembles real life	RRL
Format	F	Results	R
Fun	Fu	Scoring	Sc
General	G	Screen	Scr
Handwriting	HW	Spelling	Sp
Layout	L	Technology	Te
Legibility	Le	Time	T
Less Comfortable	LC	Tradition	Tr
Log in	Log	Typos	Ty
Manipulation	M	Writing	W
Mistakes	Mis		

Yellow themes were first identified in the students' responses concerning PPT advantages, **blue themes** were added for the PPT disadvantages, **red themes** emerged in the CBT advantages, and finally **green themes** in the CBT disadvantages. The themes are related in the analysis and not all of them appear in all the stages of the research.

Appendix 4B
Study 1:2015

FEEDBACK FORM

UCO:

Tick the mode of the test you have taken today:
COMPUTER BASED
PENCIL AND PAPER

One thing I liked about the test today:

One thing I didn't like about the test today:

Circle the answer which best describes your attitude:
A/ I prefer the PENCIL AND PAPER form.
B/ I prefer the COMPUTER BASED form.
C/ I don't mind.

Other comments:

Thank you.

Appendix 4C
Study 2: 2016

FEEDBACK FORM

Name:

One thing I liked about the PENCIL AND PAPER test today:

One thing I didn't like about the PENCIL AND PAPER test today:

One thing I liked about the COMPUTER BASED test today:

One thing I didn't like about the COMPUTER BASED test today:

Circle the answer which best describes your attitude:

A/ I prefer the PENCIL AND PAPER form.

B/ I prefer the COMPUTER BASED form.

C/ I don't mind.

Other comments:

Thank you

Appendix 5: Statistical Model – Preferences

Appendix 5A:

RESULTS: Preferences STUDY 1

Model: Destination subtest (mean subtest scores)

- The resulting model yielded an intercept of 13.2 (SE = 0.8) - this represents the average score *without* the influence of testing mode or preference, i.e. in CBT mode for people who prefer the CBT mode of testing.
- The effect of **testing mode** was not significant (beta = 0.18, SE = 0.34, p = 0.59). This would suggest no effect of testing mode (i.e. writing the test in the PPT mode did not affect the scores as opposed to writing it in the CBT).
- The effect of **preference** was not significant (beta = 0.01, SE = 0.74, p = 0.99 and beta = -0.9, SE = 0.59, p = 0.12 for “no preference” and “preference for PPT”, respectively) which suggests that preference in itself does not affect test scores (but these are test scores regardless of testing mode, so this is not surprising).
- The interaction between **testing mode** and **preference** was not significant (beta = -0.68, SE = 0.6, p = 0.26 and beta = 0.22, SE = 0.47, p = 0.64 for “no preference” and “preference for PPT”, respectively).

Model: Vocabulary subtest (mean subtest scores)

- The resulting model yielded an intercept of 15.2 (SE = 0.8) - this represents the average score *without* the influence of testing mode or preference, i.e. in the CBT mode for people who prefer the CBT mode of testing.
- The effect of **testing mode** was not significant (beta = 0.24, SE = 0.41, p = 0.55). This would suggest no effect of testing mode (i.e. writing the test in the PPT mode did not affect the scores as opposed to writing it in the CBT mode).
- The effect of **preference** was not significant (beta = -1.1, SE = 0.7, p = 0.11 and beta = -1, SE = 0.56, p = 0.05 for “no preference” and “preference for PPT”, respectively) which suggests that preference in itself does not affect test scores.
- The interaction between **testing mode** and **preference** was also not significant (beta = 0.44, SE = 0.71, p = 0.54 and beta = 0.72, SE = 0.57, p = 0.2 for “no preference” and “preference for PPT”, respectively). This suggests no effect of preference on test scores.

Appendix 5B:

RESULTS Preferences STUDY 2

Model: Destination subtest (mean subtest scores)

- The resulting model yielded an intercept of 15.9 (SE = 0.8) - this represents the average score *without* the influence of testing mode or preference, i.e. in the CBT mode for people who prefer the CBT mode of testing.
- The effect of **testing mode** was not significant (beta = -0.31, SE = 0.4, p = 0.44). This would suggest no effect of testing mode.
- The effect of **preference** was not significant (beta = 0.98, SE = 0.95, p = 0.3 and beta = 0.15, SE = 1.02, p = 0.88 for “no preference” and “preference for PPT”, respectively) which suggests that preference in itself does not affect test scores.
- The interaction between **testing mode** and **preference** was marginally significant (beta = 1.2, SE = 0.55, p = 0.032 and beta = 1.23, SE = 0.59, p = 0.038 for “no preference” and “preference for PPT”, respectively). This suggests that people who have no preference or prefer the PPT mode of testing do slightly better when writing tests in the PPT mode, unlike those who prefer CBT. However, the alpha used here is 0.05, which is rather lenient and would not survive the correction for multiple comparisons.

Model: Vocabulary subtest (mean subtest scores)

- The resulting model yielded an intercept of 14.7 (SE = 0.8) - this represents the average score *without* the influence of testing mode or preference, i.e. in the CBT mode for people who prefer the CBT mode of testing.
- The effect of **testing mode** was not significant (beta = 0.18, SE = 0.48, p = 0.7). This would suggest no effect of testing mode.
- The effect of **preference** was not significant (beta = 1, SE = 1.02, p = 0.32 and beta = -0.92, SE = 1.1, p = 0.4 for “no preference” and “preference for PPT”, respectively) which suggests that preference in itself does not affect test scores.
- The interaction between **testing mode** and **preference** was also not significant (beta = -0.21, SE = 0.66, p = 0.74 and beta = 0.74, SE = 0.71, p = 0.297 for “no preference” and “preference for PPT”, respectively). This suggests no effect of preference on test scores.