

Cite as:
Hillier, M. & Lyon, N. (2018), Writing e-Exams in pre-university college (refereed conference paper). Presented at Open Conference on Computers in Education, 25-28th June, IFIP TC3. Linz, Austria. <http://occe.2018.ocg.at/>

Writing e-Exams in pre-university college

Mathew Hillier^{1[0000-0001-9990-0479]} and Nathaniel Lyon²

¹Monash University, Australia, Mathew.Hillier@monash.edu

²Monash College, Australia, nathaniel.lyon@monashcollege.edu.au

Abstract. This study examined student's expressed strategies, habits and preferences with respect to responding to supervised text based assessments. Two trials of a computerised examination a system took place in an Australian Pre-University college in 2016 and 2017. Students in several classes studying Geography and Globalisation completed a sequence of practice and assessed work. Data was collected using pre-post surveys about their preferred writing styles, habits strategies in light of their choice to type or handwrite essay and short answer exams. Comparisons were made between those that elected to handwrite and those who chose to type the exam were conducted with several areas being significant. The performance (grades), production (word count) of the typists and hand-writers were also correlated and compared.

Keywords. e-Exams, writing strategies, student perceptions, affordances.

1 Introduction and Background

This study is part of a nationally funded project [1, 2, 3] looking at modernising supervised high stakes assessment within the Australian higher education context. This paper builds on previous e-exam trials held in Australian universities by investigating how students in a pre-tertiary pathway college context perceive the task of writing e-Exams.

We use the term 'e-Exam' (eExam) to specifically refer to a “timed, supervised, summative assessment conducted using each candidate’s own computer running a standardised operating system” [4]. This differentiates our work from those employing 'online' testing tools that take the format of a digital data collection form (e.g. Moodle quiz, Blackboard test, TCExam, QuestionMark Perception, ExamSoft). In our case we include the use of 'authentic' software applications fit for the purpose of the assessment task e.g. a fully functional office suite is provided to write reports or essays.

We will next explore the literature related to the writing of high stakes assessment tasks and in particular computerised exams before moving on to explain the process used in the study. This is followed by survey results and a discussion of the findings that draws out implications for future research and practice.

2 Literature

The idea of using computers for assessment has been around for 60 years [5], yet pen-on-paper still dominates most higher education and school examinations around the world. However, attention has recently shifted to modernising the exam room. The

'Dublin declaration' [6] features the idea of 'authentic assessment' [7] as one of its core recommendations when considering the use of computers for assessment. The declaration specifically recommends that an e-Examination must be an:

"authentic assessment that matches modern workplace practices and many student learning experiences". (p. xviii).

If we consider the modern classroom or the majority of work places, we find that computer and software technology is near ubiquitous. From Accountancy to Zoology, computers are now key tools of the trade. From keeping track of laboratory experiments through to report writing and bookkeeping we could be confident that a computer was involved. Similarly in higher education, reports, essays and communications are conducted via the typed medium. The vast majority of students in the developed world today use software tools as part of their course work and in addressing un-supervised assessment tasks. University students could hardly remember or even know what it would be like to hand-write their class work, yet we ask them to write examinations that can take up to three or more hours. The keyboard is now so commonly used that the very skill of handwriting is in decline [8] with a subsequent loss of the motor skills required to write proficiently. A previous study [9] found that hand-writers in exams beyond about 70 minutes felt physical discomfort while typists were unaffected. Fortunately there are a number of examples where the transition to e-exams is well underway or at least beginning. Prior work [10][4] described and reviewed several e-Exam projects with a longer list available [11].

As part of the conversion effort teachers and students may have an adjustment period when transitioning from hand-writing under exam conditions to typing. A survey [12] of students at the University of Bradford in the UK on their experience of using QuestionMark Perception covered a range of topics. Differences between handwriting and typing an essay style examination were explored by [13]. The study covered issues such as self-reported typing prowess, confidence, stress, use of time, writing strategies, pre-planning, structuring, editing and reviewing prior to submission. Similarly work by [14] looked at the preferences of students with regard to typing or handwriting essay exams when they were given the choice. They noted a 10% uptake by students of the typing option. Another researcher [15] reported on student's preconceptions of what an e-exam may involve prior to the start of several e-Exam trials at an Australian university with interest by students varying significantly between discipline areas. A follow up study of e-exam trials in six courses at the same institution [9] showed the typing option was selected by 5% to 34% of students. The top three comments from typists were that they could type faster, it was neater than handwriting and that they were easily able edit their work leading to more polished responses.

Finally we turn to the matter of performance. Stakeholders are concerned that computerising the exam room may have an uneven impact on student performance by advantaging some and disadvantaging others [16]. Concerns over typing prowess, speed, and computer access have been raised. However similar concerns could be raised with respect to the un-even abilities of students in terms of speed, neatness and physical strength for handwriting long exams (ibid). Performance in an e-exam in a Dentistry unit and found that while there was a moderate increase in marks for those that typed the exam compared to those handwriting. There could be no causal link

established due to the self-selecting sample groups [2]. Hand-writers that produced more words generally did better while more typed words did not see marks increase by as much. Another study [17] compared typists and hand-writers and found that the typists produced around 20% more words in an e-exam. Similarly, the study by [16] found that typists produced more words than hand-writers but there was no statistically significant difference in grades. In terms of presentation effects, [18] found no significant difference in scores due to presentation mode (typed or transcribed handwritten) on a large scale writing assessment. Prior studies have also looked at demographics such as gender [19] and how this may impact a student's performance when faced with a typed examination.

The review of prior work has outlined several areas of focus that informed questions to be investigated (see Table 1) and an up-front frame for analysis.

Table 1: Areas for investigation

Area	Example research questions
Rationale of students	What proportion of candidates were in favour of typing their exam? What rationale was provided for their choice? Did the e-Exam environment support their writing?
Writing strategies	Were there differences in the writing preferences and strategies used by those that typed and those that handwrote?
Student performance	Were there differences in words produced and grades achieved by those that typed and those that handwrote?

2 Study Context

In Australia, the first author is leading an Australian federal government funded grant project [1] investigating how authentic e-assessment can be introduced into the examination rooms of universities using BYOD. This study was undertaken within two units offered in the Foundation Year [20] program at Monash College, Australia. The program is at an equivalent level as an Australian year 12 high school certificate or the International Baccalaureate. The second author, is a unit coordinator and teacher in the two units in which the e-exam trials were conducted. The trials were carried out using in-class supervised written assessments requiring short text and essay style responses.

3 Method and Approach

Students in several classes undertaking Geography (Geo) in semester 1, 2016 and Globalisation (Glo) in semester 2 2017 at Monash College took part in a two phase trial of the e-Exam system. Ethics protocol approval was gained via Monash University prior to the trials. In this study a fully functional word processor was used that was part of the e-Exam platform as described by [3]. Students used a custom Live Linux USB environment on their own laptops that provided a full office suite. They could use editing tools such as spelling, grammar and highlighting to help with their

writing tasks. Being able to copy, delete or move text around with ease means that there are functional differences to undertaking the same task using pen-on-paper.

In phase one, students participated in an in-class, ungraded preview session. They were provided instructions on how to use the e-Exam system with their laptops and the opportunity to practice the e-exam processes, use the software and try the question response format. Students completed a pre-exam survey that requested technical information about their laptop and their first impressions of the e-Exam process and software. Attitudes and responses to use of the e-Exam software are reported separately in [21].

Phase two occurred two weeks later in the form of an in-class, graded, supervised assessment task for which students could choose to type or handwrite. Materials were provided on paper or as a word processor document and included photos, diagrams, charts and data tables. In Geo the assessment task was a single case study with an extended essay response. In Glo two short answer sections and a mini-case essay response were required. Students then completed a post-exam survey before leaving the room. An extended account of the trial process is provided in [21].

Qualitative survey data relating to student's opinions on writing in exams was analysed using SPSS v24. Likert items were treated as non-parametric as per advice by [22]. This stance is supported by [12] when analysing students' perceptions of e-Assessment. The Mann & Whitney's U test [23] was used to test the variance between groups (males v females and typists v hand-writers) of Likert responses. Chi-square was also used to test if gender played a role in the decision to type the exam. A Fisher's exact test [23] was used when comparing categorical variables. When it came to performance data (grades) a T-test was used to compare between groups while a Spearman's test [25] was used to test for correlation between word count and grades. We used an alpha level of .05 for all statistical tests, unless otherwise noted. However the participants were not randomly assigned to the typing or hand-writing groups which makes the results only descriptive of this study. In terms of items reflecting student's opinions we used statistical tests as a tool to summarise rather than to be representative of an objective truth as per [12].

4 Findings

The trials involved 128 pre-tertiary students, 65% were female and 35% were male. In the exam, 52% of students elected to type. Table 1 shows participation at each stage.

Table 2: Intention to type the exam at each stage of the study (counts)

Participation	Initial interest	After practice	At the exam
Type	73	72	64
Handwrite	26	30	59
Missing	29	26	5

A Chi-square test revealed a statistically significant difference in relation to gender and choice to type the exam ($\chi^2(1) = 5.299, p = 0.021$) with 68% of males choosing to type compared to 46% of females. However the differences due to gender were not

significant in the earlier stages of the trial. This may be due to the larger number of undecided ('missing') cases in the earlier stages.

Opinions

We examined the differences in opinions on five point Likert items (where strongly agree 5, neutral 3 and strongly disagree 1) between those that typed the exam and those that handwrote it using a Mann-Whitney U test. Table 3 displays the extent of agreement with means and standard deviations provided for clarity along with the results of significance tests with the difference between the means shown.

Table 3: Post-exam survey responses by text production mode

Question	Typed			Handwrote			diff	MW	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		<i>U</i>	<i>p</i>
Post Exam Survey									
I type faster than I handwrite	64	4.1	1.1	55	1.8	1.2	2.2	412	>.001
I type accurately	63	4.1	0.8	52	2.5	1.1	1.7	379	>.001
When I make errors, I am able to quickly correct them when typing	64	4.3	0.8	53	2.9	1.3	1.4	634	>.001
I often rely on spell check to detect errors	62	3.8	1.3	54	2.9	1.3	0.9	1060.5	>.001
I work more efficiently when I type on a familiar keyboard	64	4.2	0.9	53	3.0	1.3	1.2	721.5	>.001
My hand-writing is normally neat and legible	63	3.6	1.3	52	3.1	1.1	0.5	1183.5	0.008
I go back to re-read and revise my writing quite a lot	63	3.8	1.0	53	2.9	1.2	0.9	969.5	>.001
I prepare most of my assignments /reports using a computer	64	4.0	0.9	53	3.0	1.4	1.0	1012.5	>.001

Those that chose to hand-write the exam were asked about the neatness of their writing and comfort levels experienced during the exam (see Table 4). The exam duration was 70 minutes in both cases.

Table 4: Post-exam survey responses on hand-writing

Question	<i>n</i>	<i>M</i>	<i>SD</i>
I think my hand writing was neat and legible	53	3.6	0.9
I experienced discomfort (sore/tired/cramp) in my writing hand	53	2.9	1.2

Typists in the Geography unit were also asked to reflect specifically on using a computer for the exam given the nature of the task (See Table 5).

Table 5: Post-exam survey responses on using a computer for the assessment

Question	<i>n</i>	<i>M</i>	<i>SD</i>
I was able to produce a better final version of this assessment	24	4.3	0.6
I was able to quickly complete the assessment	24	4.3	0.7
I was able to easily edit and make changes	24	4.6	0.6
I was able to easily refer to reference materials and resources	24	3.7	1.0
I was able to easily think and compose my answer using a computer	24	4.3	0.6
I would like to use a computer for similar assessments in the future	23	4.2	0.7

Students were asked about their preferences for production method ('computer', 'same' or 'pen and paper') for a range of writing activities, style and features. A Fisher's exact test for categorical variables was used to see if their preferences may have influenced their choice to type or handwrite the exam (see Table 6). Counts and percentages in brackets are shown for each production method (C, S or P) for the two groups (typists and hand-writers). Significant differences were noted across many of the items with preferences in alignment with their actual choice of exam mode.

Table 6: Post-exam survey responses on exam writing strategies

Question	Typed				Handwrote				Fisher <i>p</i>
	C	S	P	<i>n</i>	C	S	P	<i>n</i>	
C=computer, S=same, P=pen									
I write more words in an exam when...	35 (56)	23 (37)	5 (8)	63 (53)	1 (2)	18 (32)	37 (66)	56 (47)	>.001
I write faster in an exam when...	42 (66)	17 (27)	5 (8)	64 (53)	3 (5)	9 (16)	44 (79)	56 (47)	>.001
I think more carefully before I start writing in an exam when...	32 (50)	13 (20)	19 (30)	64 (54)	3 (5)	17 (31)	35 (64)	55 (46)	>.001
I pause to think most in an exam when...	21 (33)	19 (30)	24 (38)	64 (53)	11 (20)	26 (46)	19 (34)	56 (47)	0.123
I write in a style that feels more normal in an exam when...	29 (46)	24 (38)	10 (16)	63 (53)	1 (2)	17 (30)	38 (68)	56 (47)	>.001
I try not to make changes unless they are really important when...	18 (28)	18 (28)	28 (44)	64 (53)	3 (5)	25 (45)	28 (50)	56 (47)	0.003
I change, move or correct words or phrases most when...	42 (67)	13 (21)	8 (13)	63 (53)	11 (20)	21 (38)	24 (43)	56 (47)	>.001
I think the overall structure/argument of my response is better when...	38 (59)	15 (23)	11 (17)	64 (53)	3 (5)	21 (38)	32 (57)	56 (47)	>.001
I make more effective use of the time available in an exam when...	43 (67)	16 (25)	5 (8)	64 (53)	2 (4)	15 (27)	39 (70)	56 (47)	>.001
I go back and read over my response most in an exam when...	28 (44)	28 (44)	8 (13)	64 (53)	3 (5)	23 (41)	30 (54)	56 (47)	>.001
I feel more stressed in an exam when...	15 (23)	28 (44)	21 (33)	64 (53)	18 (32)	22 (39)	16 (29)	56 (47)	0.589
I am more likely to run out of time in an exam when...	13 (20)	28 (44)	23 (36)	64 (53)	16 (29)	22 (39)	18 (32)	56 (47)	0.601
Overall I feel I perform better in an exam when...	37 (59)	22 (35)	4 (6)	63 (53)	1 (2)	14 (25)	41 (73)	56 (47)	>.001

Performance

Performance was expressed as a percentage grade. There was a statistically significant performance difference between the two units (Geo, $n=38$, Mdn 73.75 and Glo, $n=85$, Mdn 57.14) using a Mann-Whitney U test ($U=1100$, $p=.005$).

When the grade data was grouped by gender as (shown in Table 7) it was not found to be statistically significant, although the result was borderline at $p=.050$. In the Geography unit a significant difference in grades between genders was found ($p=.006$) however this was not so in Globalisation. Normal distributions were confirmed for each gender pair using standardised skewness and the Shapiro-Wilks

test with Levene's test (shown) demonstrating equivalence of variance. Means and standard deviations are also provided for clarity.

Table 7: Grades out of 100 by gender

Unit	Female			Male			Df	T-test		Levene's test	
	n	M	SD	n	M	SD		T	P	F	p
Geo	24	76.2	17	12	56.4	23	(1,34)	2.912	0.006	1.422	0.241
Glo	53	58.3	18.5	29	56	20.7	(1,80)	0.519	0.605	0.002	0.961
Both	77	63.9	19.8	41	56.1	21.1	(1,116)	1.98	0.050	0.005	0.943

When comparing the grades of typists and hand-writers (shown in Table 8) across both units, no statistically significant difference was found. However within the Globalisation unit there was a small statistically significant difference in grades between typists and hand-writers ($p=.033$). The distribution of each pair was found to be normal using standardised skewness and the Shapiro-Wilks tests, with Levene's test (shown) establishing equivalence of variance. Means and standard deviations are provided for clarity.

Table 8: Grades out of 100 for typists and hand-writers

Unit	Typed			Handwrote			Df	T-test		Levene's test	
	n	M	SD	n	M	SD		T	p	F	p
Geo	25	64.3	21.5	13	76.7	18.8	(1,36)	-1.754	0.088	0.212	0.648
Glo	39	62.9	21.0	44	53.7	17.4	(1,81)	2.174	0.033	1.636	0.205
Both	64	63.4	21.0	57	59.0	20.1	(1,119)	1.190	0.236	0.333	0.565

Within each of the Geography group ($n = 38$) and the Globalisation group ($n = 32$) positive, statistically significant correlations were found between the number of words written and the grade achieved from Spearman's tests (Geo: $r_s = .628$, $p = >.001$ and Glo $r_s = .865$, $p = >.001$).

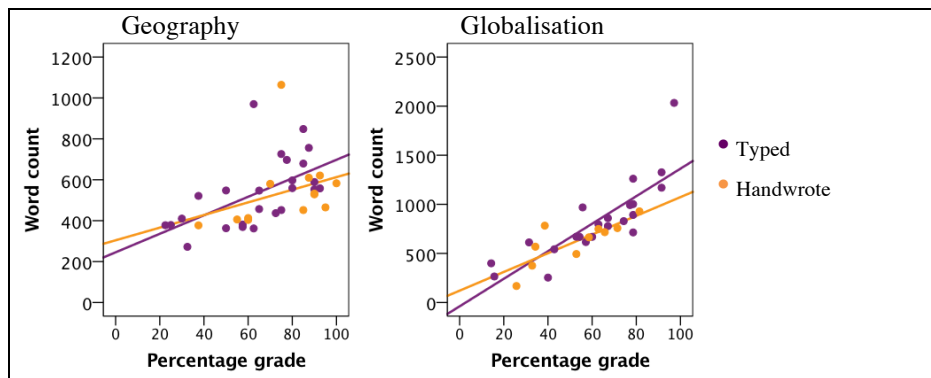


Figure 1: correlation between number of words and grade

Mann-Whitney U tests found no statistically significant difference in the number of words produced by typists and hand-writers (see Table 9). It is worth noting that word count data was only available for half the Globalisation group and that assessment tasks were different between Glo and Geo so the two groups are not comparable.

Table 9: words produced by typists and hand-writers

Unit	Typed			Handwrote			diff	MW	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		<i>U</i>	<i>p</i>
Geography	25	536	170	13	541	178	5	157.5	0.878
Globalisation	22	833	83	10	620	71	213	71	0.113

5 Discussion

The choice to type or handwrite appears to be multifaceted but with the strongest differences declared around being able to type more quickly than handwrite and typing accuracy. Overall the results showed an alignment between writing strategies and choice of text production mode in the exam (see Table 3 and 6). The result of 52% of students electing to type is much higher than reported in previous studies involving free choice [9, 14]. The use of in-class practice sessions compared to optional out-of-class practice sessions used by [9] may have contributed to the increased uptake of the typing option in our study.

In both 70 minute assessment tasks only a minority of hand-writers experienced discomfort (i.e. agreement of 2.9 out of a maximum 5 with standard deviation 1.2 indicated that only some were impacted). This was consistent with previous work [9] where a 70 minute duration was the cross-over point where discomfort started to become a problem.

Typists generally agreed that the computer allowed them to produce better responses on the assessment task (i.e. most Likert items in Table 5 were rated above 4 on the 5 point scale), including that they would like to be able to use a computer for similar assessments in the future. The weakest agreement related to their ability to easily refer to reference materials. In this study students had to scroll up and down between the materials and their written response. Providing reference materials in a separate file (that would allow side by side window arrangement or 'Alt-Tab' between windows), in a split screen format within the document [26] or on paper may help in this regard.

In terms of performance there were mixed results. The number of words produced by typists was greater in Glo, as per [16, 17], but less in Geo. The classes are relatively small and assessment tasks were different so this may play a role. In Geo hand-writers achieved slightly better grades (but not significantly so) and in Glo typists did better with a borderline statistically significant difference in grades. When comparing by gender, in Geo females did significantly better. In Glo they also did slightly better but not significantly so. Overall a weak link between gender and performance could be claimed within these groups. In both classes we found a statically significant correlation between words produced and marks. The scatter plot in figure 1 shows that the correlation occurred over a narrow band in Geo with a minimum of 300 words required before a pass was possible. Yet an adequate word count certainly does not guarantee a pass. Similarly a larger number of words above 600 did not see the highest grades, indeed those with the highest grades wrote about the same as the group mean of 538 words. The scatter plot for Glo shows a more tightly monotonic relationship between words and grades. Again, correlation is certainly not causation in terms of word count – quality still matters!

6 Conclusion

Overall we observed a significant alignment between preferred writing strategies and choice of text production method in the two trials. Further, the grades achieved between typists and hand writers did not differ significantly. These two facts are not surprising when dealing with thinking, purposeful humans who aim to maximise outcomes. However this does raise the concern that should a shift occur towards fully typed assessment then a period of transition with assistance for those who preferred handwriting should occur to ensure they are able to adapt successfully.

Acknowledgments. The authors would like to thank the Australian Government Office of Learning and Teaching for financial support and the students at Monash College for being willing to 'give it a go'.

References

1. TEAA (2015) Transforming exams across Australia, Australian Government Office for Learning and Teaching, Grant, ID15-4747 <http://transformingexams.com>
2. Fluck A. & Hillier, M. (2016): Innovative assessment with eExams. Australian Council for Computers in Education Conference, Brisbane, 29 Sep to 2 Oct. Retrieved from <http://conference.acce.edu.au/index.php/acce/acce2016/paper/view/34/27>
3. Hillier, M. & Fluck, A. (2017). Transforming Exams - How IT works for BYOD e-Exams. Australasian Society for Computers in Learning in Tertiary Education conference (pp. 100-105). Toowoomba, 4-6 December. <http://2017conference.ascilite.org/wp-content/uploads/2017/11/Concise-HILLIER.pdf>
4. Fluck, A. & Hillier, M. (2017). eExams: Strength in Diversity. In A. Tatnall & M. Webb (Eds.), *Tomorrow's Learning: Involving Everyone. Learning with and about Technologies and Computing* (Vol. 515, pp. 409–417). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-74310-3_42
5. Swets, J. A. & Feurzeig, W. (1965). Computer-Aided Instruction. *Science*, 150(3696), 572–576. <http://doi.org/10.1126/science.150.3696.572>
6. Tatnall, A., & Webb, M. (2017). Tomorrow's Learning: Involving Everyone. Learning with and about Technologies and Computing: 11th IFIP TC 3 World Conference on Computers in Education, Dublin, Ireland, 3-6 July, Revised Selected Papers. Springer
7. Crisp, G. (2009). Towards Authentic e-Assessment Tasks (Vol. 2009, pp. 1585–1590). EdMedia: World Conference on Educational Media and Technology, Honolulu, HI, USA: Association for the Advancement of Computing in Education. Retrieved from <http://www.editlib.org/p/31689/>
8. Sülzenbrück, S., Hegele, M., Rinkenauer, G., & Heuer, H. (2011). The Death of Handwriting: Secondary Effects of Frequent Computer Use on Basic Motor Skills. *Journal of Motor Behavior*, 43(3), 247–251. <https://doi.org/10.1080/00222895.2011.571727>
9. Hillier, M. (2015). To type or handwrite: student's experience across six e-Exam trials. Proceedings of the Australasian Society for Computers in Learning in Tertiary Education conference, Perth, Australia: ASCILITE. Retrieved from http://transformingexams.com/files/Hillier_2015_ascilite_fp.pdf
10. Fluck, A., Pålsson, H., Coleman, M., Hillier, M., Schneider, D., Frankl, G., & Uolia, K. (2017). eExam symposium: design decisions and implementation experience. Presented at the IFIP World Conference on Computers in Education, 3-6 July, Dublin, Ireland. Retrieved from http://transformingexams.com/files/Fluck_etal_2017.pdf

11. Hörnblad, P., Brenner, M. (2016) Digital Exam. SUNET Incubator project, NORDUnet. <https://portal.nordu.net/display/Inkubator/Digital+Tentamen>
12. Dermo, J. (2009) E-assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40, 203–214
13. Mogey, N., & Fluck, A. (2014). Factors influencing student preference when comparing handwriting and typing for essay style examinations: Essay exams on computer. *British Journal of Educational Technology*, n/a-n/a. <https://doi.org/10.1111/bjet.12171>
14. Purcell, M., Paterson, J., & Mogey, N. (2012). Exams: Comparing Handwritten Essays with those Composed on Keyboards (Final Report). Heslington, UK: The Higher Education Academy. <http://www.docs.hss.ed.ac.uk/divinity/About%20the%20school/Elearning/HEA2012-Final%20Report.pdf>
15. Hillier, M. (2014). The Very Idea of e-Exams: Student (Pre)conceptions. Proceedings of the Australasian Society for Computers in Learning in Tertiary Education conference, Dunedin, New Zealand. Retrieved from <http://ascilite.org/conferences/dunedin2014/files/fullpapers/91-Hillier.pdf>
16. Mogey, N., Paterson, J., Burk, J., & Purcell, M. (2010). Typing compared with handwriting for essay examinations at university: letting the students choose. *Research in Learning Technology*, 18(1). <https://doi.org/10.3402/rlt.v18i1.10675>
17. Fluck, A. (2011) eExaminations Strategic Project Final Report for Academic Senate, (unpublished report), University of Tasmania. Meeting 1/2011.
18. Rankin, A. (2015). A comparability study on differences between scores of handwritten and typed responses on a large-scale writing assessment (unpublished PhD dissertation). University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/1895>
19. Terzis, V., & Economides, A. A. (2011). Computer based assessment: Gender differences in perceptions and acceptance. *Computers in Human Behavior*, 27(6), 2108–2122. <https://doi.org/10.1016/j.chb.2011.06.005>
20. Foundation Year (2016), Monash College, Australia, <https://web.archive.org/web/20160402024339/http://www.monashcollege.edu.au:80/courses/foundation-year>
21. Hillier, M & Lyon, N (2018) Student experiences with bring your own laptop e-Exams in pre-university college, Working paper, Monash University.
22. Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38(12), 1217–1218. <https://doi.org/10.1111/j.1365-2929.2004.02012.x>
23. Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>
24. Freeman, G. H., Halton, J. H. (1951). Note on an exact treatment of contingency tables, goodness of fit and other problems of significance. *Biometrika* 38: 141 – 149.
25. Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*. 15: 72–101. doi:10.2307/1412159
26. Al Nadabi, Z. (2015). Features of an online English language testing interface. Australasian Society for Computers in Learning in Tertiary Education conference (pp. 369–373). 30 Nov - 3 Dec, Perth, Australia <http://www.2015conference.ascilite.org/wp-content/uploads/2015/11/ascilite-2015-proceedings.pdf>