

Effect of typeface characteristics on student's performance - A case study for design aptitude tests

Kaveri Sharma, MIT Institute of Design, Pune, kaverisharma09@gmail.com Anudnya Marathe, MIT Institute of Design, Pune, anudnya.marathe@mitid.edu.in Ranjeet Waikar, MIT Institute of Design, Pune, ranjeet.waikar@mitid.edu.in Debayan Dhar, MIT Institute of Design, Pune, debayan.dhar@mitid.edu.in

Abstract: Competitive exams like design entrance test requires student assessment based on multiple aptitude constructs such as analytical reasoning and analogical skills. The representation pattern of the questions for each of these aptitude constructs might play a major role in information processing behavior of the students in such assessment environments. The present study investigates the effect of item representation (using different typefaces) on the students' performance in a design aptitude test environment and reports the effect of typeface variations of test items on cognitive load and performance of students in a speed test environment. The results highlight that variation in typeface characteristics of test items affects cognitive workload and performance of students in a design aptitude test. The results of the study has immediate effect for test formulators of design aptitude tests and would provide them with insights for designing an engaging exam experience for the students.

Key words: Design Aptitude, Typeface Characteristics, Test typography, Cognitive Load.

1. Introduction

Tests in schools and colleges are generally using a conventional medium to assess a student's capabilities and learnability. In an assessment environment the candidates are under an immense amount of stress and anxiety which affects performance and hence compromises the construct validity of examination. A study done by Hembree (1988) suggests that high level test anxiety causes decrements in cognitive performance. An area which has been of interest to the design research community is design of content and

presentation with the objective of minimizing any variance in performance which might arise due to factors which are beyond candidate's control. The assumption for this study is that an effective medium of communication can be one which integrates visual and text seamlessly to ensure a better comprehension. Willi Kunz in the book titled 'Typography -Macro and Micro aesthetics' presses on the importance of typography in integrating and balancing form and function, he goes on to say that typographic form and message content are inextricably linked and design should not only objectively convey information but also give subjective cues for interpretation of content.

The visual expression of typeface form contributes towards readability and legibility of the text and hence has a major to play in deciding appropriateness of a typeface for the context. The intent was to explore effect typographic form variation might have on test performance. In a speed test environment where students are required to direct the cognitive abilities towards analysing the problem rather than understanding text the form can be visually manipulated to represent content. Although the form variation may result in an interruption in flow, the interruption can be used as a tool to direct the candidate's attention towards important part of instruction or question.

This paper reports the result of an experimental study in which 39 students participated, result indicate that there is no significant difference in test score when presentation of the question is changed through variation of typographic characteristics of the instruction text. However there is a significant variation in the lower bound of mental demand for one of the variations which leads us to conclude that any additional mental processing effort which arises in an assessment environment due to effect of presentation does not significantly affect the score of the candidate.

2. Literature Review

Tests plays major role in educational area. Typeface, also known as font, is made up of all the characters of one particular design, regardless of size (Chelesnik, 2009). Every possible effort needs to be made to create the format using these typefaces to improve the usefulness. The impact of test typography on students' achievement, levels of cognitive load, and typographic design and pattern preferences has rarely been studied. "The successful performance on tests often depends on students' ability to read, decode, comprehend, and respond to written text" (Hanson, Hayes, Schriver, LeMahieu, and Brown, 1998, p.2) This study aims to answer this question by utilizing research in the field of variations in Typography design to make decisions about how to best format a test. Role of typography in design entrance exam has not been studied from the point of view to check how the variation in typography of design questions affects the cognitive load of

students while solving paper base exam (Chelesnik, 2009). There is however, review of literature based upon the cognitive load of variations in typography, semantic aspects of typography and variations in element form and space in typography.

Cognitive load has been measured by a learner's difficulty level. In addition, a learning efficiency metric has been used to quantify the efficiency of instruction (e.g., Clark, Nguyen & Sweller, 2005; Tuovine & Paas, 2004). The efficiency score is calculated by performance and mental effort, meaning a high efficiency condition occurs when performance is higher and effort is less (Tuovinen & Paas, 2004). Based on cognitive load theory, researchers and practitioners have proposed interface design guidelines (e.g., Norman, 1998; Shneiderman & Plaisant, 2005; Swan, 2001). Principles of learning that apply to the design and development of documentation includes cognitive load and constructivism. Cognitive load is concerned with long-term memory, working memory, and contextual relevance. Cognitive load is about balancing the amount of information, structuring the delivery into manageable chunks, and maintaining content relevance for the learner (Sweller & Chandler, 1994).

The few thorough studies carried out by research teams (e.g. Burnhill, Hartley, Frase & Young, 1975; Hartley & Burnhill, 1976; Hartley & Trueman, 1981; Spencer, Reynolds & Coe, 1974, 1975) on the structure and articulation of information on the page have provided useful findings for the design of written information other than examinations. As with these past studies, the aim of the present study is to explore how the layout of text as a whole (i.e. the combination of various typographic features) affects performance. Testing three different variations for presenting typography in this way reflects the situation where candidates are presented with three text layouts per examination, which differ in various typographic features (not just in one feature). Variation of typography in terms of semiotic and elements of form space and structure has not been studied yet.(e.g. Hartley & Burnhill, 1976; Lund, 1999) (Maria dos Santos Lonsdale, Mary C. Dyson and Linda Reynolds, 2006)

In addition to standing in for spoken language, the semantics of typography resides in its wealth of connotative meanings derived from graphic detail and in its pictorial qualities (StöckI, 2005). Typography can, and is, used ideationally, to represent actions and qualities. (Leeuwen, 2006) Type size is defined as the height of typeface characters. It is the distance from ascender line to descender line and is measured in points. (Chelesnik, 2009) Furthermore, various typefaces that are considered 12-point may not be the exact same size because some typefaces use different measurement systems (Hartley, 1987) White space is the amount of inked surface relative to blank paper. Appropriate use of white space is a powerful design element. It makes text easier to read because it provides spots where readers can rest their eyes and gives the brain a chance to process information (Baker, 2001; Bradshaw & Johari, 2002; Cubberley, 1991). Legibility is enhanced by manipulating typeface, white space and the overall simplicity of design (Gribas & Sykes, 1996). Typeface selection determines the appropriate amount of line spacing (Bix, 2002). As pointed out by Lupton (2003), the interaction of typographic features is complex, and changing one feature affects each of the others. First, typography works on the 'ideational' level as it refers to, comments on or reinforces verbal messages of the text - pictorial typography can express ideas on its own by virtue of representing objects. Second, typography functions 'inter-personally': it says something about the nature or emotional state of the writer, anticipates the aesthetic inclinations of the addressees or indicates the nature of the communicative contact between writer and reader. Third, graphic design is 'textual' when it serves to visually structure a verbal message and bring out its logical make-up. Clearly, typography can have numerous functions and it would be a rewarding venture to study the systematic relations between the use of (typo)-graphic means and their communicative functions with respect to text and linguistic structures (Stöckl, 2005).

Figure 1 lists factors which affect the subjective performance of candidates in assessment environment.



Figure.1 Attributes of typography affecting the subjective performance of candidates in assessment environment

Based on the literature review, following research questions may be raised regarding design aptitude test performance-

- 1. Will there be any affect on the student's performance after changing the typography presentation style?
- 2. In test papers typography, if the implementation of EFSS (Elements of form space and structure) or semantics principles happens, will this affect the cognitive load of a student?

2.1 Hypothesis

"Change in representational features of test items would affect student's performance in Design Aptitude Test.

Figure. 2 elaborates probable model of effect of semantic and EFSS typographic questions presentation styles for design aptitude test.



Figure.2 Schematic diagram of Design Aptitude Test Environment

Working Hypotheses

- H_{a1}: "Representational features of test items in terms of typeface characteristics affect performance of students"
- H_{a2}: "Representational features of test items in terms of typeface characteristics affect cognitive load"

3. Method

The study investigates the hypothesis in context of a design aptitude test; hence, two constructs contributing to design aptitude were shortlisted. A screening test was conducted to select students with similar level of problem solving abilities for final experiment. The sample for screening test consisted of first year undergraduate students of MIT ID, Pune. 120 students were asked to take a thirty minute test comprising of 10 multiple choice questions on analogical thinking and analytical reasoning. Test was paper based to avoid any test mode effect. All the questions required the candidate to give answer by marking correct option on original test sheet. The questions were selected after a review by a subject matter expert to ensure that complexity level remains same. A purposive sampling procedure was adapted and 51 students were shortlisted for final experiment based on their scores in screening test.

Shortlisted students were allotted to three different groups for final experiment using the process of random sampling. Since primary focus of this study was on typographic forms, we decided to create three versions of test with different typographic variations to assess its effect on student's performance. The basic layout for the entire test was kept the same across all versions. Out of the 51 shortlisted students, 39 students appeared for final experiment. The percentage of females to males was 66.67% to 33.33%. Age range of participants was from 17 to 21.

Version 1: The control group was administered a test which was similar to a conventional paper based test with no variation in form of typeface. A sans serif Calibri typeface was used for this set (Figure. 3).

Which is the odd one out?
a) Congregation
b) Dispersion
c) Compilation
d) Convocation
e) Aggregation

Figure.3 Example of question items in version 1

Version 2: The second version comprised of variation in form on basis of semantics. The focus was on conveying the meaning of content through visual representation of type form to increase the communicative quality of the message (Figure. 4).

```
which is the odd one out? a. Congregation
b. Dispersion
c. Compilation
d. Convocation
e. Aggregation
```

Figure.4 Example of questions items in set 2

Version 3: Third version consisted of variation on the basis of interplay between space and structure of typographic forms (Figure. 5).

a. Congregation b. Dispersion c. Compilation d. Convocation e. Aggregation

Figure.5 Example of questions items in set 3

After completion of 30 minute test, participants were asked to fill up NASA TLX questionnaire to measure the subjective work load and the underlying dimensions of workload during the assessment process.

4. Results

4.1 Test Score

To understand the effect of typographic variation (in test questions representation pattern) on student's score test score of different groups were analyzed. Mean test scores of each group were compared using Mann Whitney U Test.

There are mean differences observed in test scores due to question representation format. For the conventional question representation with Calibri, mean test score was 6.58. Similarly, for question representation with sematic typeface and EFSS, mean test scores were 6.66 and 6.06 respectively. Mean value of test score was higher for question representation format with sematic typeface (version-2) than the conventional typeface format (version-1) and the EFSS typeface format (version-3) (See **Figure 6**). Pairwise comparison (employing Mann-Whitney U-test) of obtained test scores among three experimental groups (conventional, semantic and EFSS) do not show any significant differences between test versions (See **Table 1**). Though significant differences were not observed in mean test scores, but, it was observed from calculated effect sizes (*r*-values) for all comparisons that there is a tendency to differ mean values of test scores between test groups (version 1 vs. version 2, version 1 vs. version 3 and version 2 vs version 3).



Figure.6 Error bar of Test Score for different versions of the test

	Version 1 vs Version 2	Version 1 vs Version 3	Version 2 vs Version 3
Mann-Whitney U	70.00	67.00	67.50
Z	119	-1.164	-1.129
Asymp. Sig. (2-tailed)	0.905	0.244	0.259
Effect size (r)	0.02	0.22	0.21

Table 1. Comparative Results of Test score using Mann Whitney U Test

4.2 Cognitive Load

Cognitive workload analysis was done using NASA-TLX scale. There were no significant differences in overall NASA-TLX scores among three experimental groups when pairwise comparisons were done in different combination applying Mann-Whitney U test (See Figure 7, Table 2, Table 3, and Table 4). Average overall NASA-TLX score was higher in case of version-1 ($M_1 = 42.00$, $SD_1 = 11.47$) than version- 2 ($M_2 = 41.25$, $SD_2 = 9.73$) and version-3 ($M_3 = 41.46$, $SD_3 = 11.88$). Though there are mean differences observed in scores of individual NASA-TLX scores were also compared among three experimental groups, but, no significant differences were found (See Figure 8, Table 2, Table 3, and Table 4).



Figure.7 Error bar of NASA TLX Overall Score



Figure.8 Error bar of different subscales of NASA TLX

			· · · · ·			
Table 2	NASA TI X Score	for Control Grou	in (Test Version	1) vs Fyn	periment Grour	n 1(Test Version 2)
				1) V3 LAP		

	Overall	Mental	Physical	Temporal			
	Rating	Demand	Demand	Demand	Performance	Effort	Frustration
Mann-Whitney U	66.00	67.50	61.00	51.00	62.00	64.50	61.00
Z value	348	262	683	-1.222	583	437	643
Asymp. Sig. (2-tailed)	.728	.793	.495	.222	.560	.662	.520

Table 3. NASA TLX Score for Control Group (Test Version 1) vs Experiment Group 2 (Test Version 3)

	Overall	Mental	Physical	Temporal			
	Rating	Demand	Demand	Demand	Performance	Effort	Frustration
Mann-Whitney U	87.00	90.00	68.00	69.00	84.50	75.00	79.00
Z value	147	.000	-1.137	-1.036	270	740	550
Asymp. Sig. (2-	.883	1.000	.256	.300	.787	.459	.583
tailed)							

Table 4. NASA TLX Score for Experiment Group 2 (Test Version 2) vs Experiment Group 3 (Test Version 3)

	Overall	Mental	Physical	Temporal			
	Rating	Demand	Demand	Demand	Performance	Effort	Frustration
Mann-Whitney U	83.00	86.50	81.50	88.50	82.00	69.00	59.00
Z value	342	173	433	073	394	-1.033	-1.530
Asymp. Sig. (2-	.732	.863	.665	.941	.693	.302	.126
tailed)							

5. Discussion and Conclusion

From the results it was observed that obtained test scores of participants were higher in case of test mode version-2 (question presentation with semantic type face) than the conventional test mode whereas cognitive load was comparatively lower among participants in case of version-2 than the conventional test mode. Therefore, it may be

concluded that questionnaire presentation through semantic typeface is better for design aptitude test than the conventional question format. Though the mean differences was not significant in test score comparisons due to small sample size, but, there is a tendency to get significant mean differences as evident by comparatively strong effect size (r) in between group comparisons (version 1 vs. version 3 and version 2 vs version 3). Therefore, further study can be done to get significant differences between groups with large sample size.

There have been studies on typography in a test based environment in the past which have focused on attributes other than form of typeface; Nibblelink (1993) and Gerig (1988) concluded that type size does not affect student achievement on math test; similar results were reported for science tests by Chelesnik (2009). Results of the present paper have consistency with the past research that variation in typographic characteristics do not have any significant effect on the test score among candidates participated in different groups. Although the results of this study indicate that the lower bound of mental demand is higher for test version 3, its effect is not significantly visible in the mean score of the candidates which leads us to conclude that the presentation might not be of importance when the level of mental processing is high. The result of the study also directly support the learner centered design principles [as reported by Sweller, (1994)] and would provide guidelines for instructional designers for designing educational assessment systems. Limitation of the present investigation includes:

- This study did not take into account the time taken by the candidates to complete the different versions of the test, whether the type form variations are impacting comprehension time and hence the time taken by students to answer question can be investigated in future.
- Preference of candidate for a particular version.

Therefore, these issues can be considered as future scopes of the present research.

6. Acknowledgement

We would like to express our sincere gratitude to Prof. Anirban Chowdhury for his insightful advice and contribution which helped us in completion of this study.

References

- Baker, W. (2001). HATS: A design procedure for routine business documents, Business Communication Quarterly - ERIC
- Bix, L. (2002). The elements of text and message design and their impact on message legibility: A literature review, Journal of Design Communication
- Bradshaw, A. C., & Johari, A. (2002). Effects of white space in learning via the web, Journal of Educational Computing Research, 2002
- Chelesnik, A. (2009). The Impact of Test Typography on Student Achievement, Anxiety, and Typographic Preferences. California State University San Marcos
- Cheon, J. and Grant, M. (2007). Do flashy interfaces help? Comparing learning efficiency of different interface types based on cognitive load theory
- Cubberley, C. W. (1991). Write procedures that work, Library journal ERIC

Dutke, S. and Stöber, J. (2001). Test Anxiety, Working Memory, and Cognitive Performance: Supportive Effects of Sequential Demands

- Gerig, J. A. (1988), The effects of adjusting print size on reading comprehension and mathematical problem-solving performance, *Dissertation Abstracts International*, *50* (02), 356A.
- Gribas, C., & Sykes, L. (1996). Creating great overheads with computers, College Teaching Taylor & Francis
- Hanson, M.R., Hayes, J.R., Schriver, K., LeMahieu, P.G., and Brown, P.J. (1998). A plain language approach to the revision of test items. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA, April 16, 1998
- Hartley, J. (1987). Typography and executive control processes in reading. In B. K. Briton & S.M. Glynn (Eds.), Executive control processes in reading (pp. 57-79). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. Review of Educational Research, 58, 47-77.
- Hogan, D.B. (2013). THEORIES THAT APPLY TO TECHNICAL DOCUMENTATION.
 - Connexions-international professional communication journal, 155-165, ISSN 2325-6044
- L Šilić, J Dolić, J Pibernik (2009) Connotative meaning of type, 3th International conference of printing, design
- Leeuwen, T.V. (2006). Towards a semiotics of typography, Information Design Journal
- Lonsdale, M.D.S. and Dyson, M. C. and Reynold, L. (2006). Reading in examination-type situations: the effects of text layout on performance
- Lupton, E. (2003). An exact science.
- Nguyen, F. and Clark, R.C. (2005). Efficiency in e-Learning: Proven Instructional Methods for Faster, Better, Online Learning
- Nibbelink, W. H., Gerig, J. A., & Hoover, H. D. (1993), The effect of print size on achievement in mathematics problem solving. *School Science and Mathmatics*, *93*(1), 20-23
- Stöckl, H. (2005). Typography: body and dress of a text a signing mode between language and image
- Sweller, J. (1994). Cognitive Load Theory, Learning Difficulty, and Instructional Design
- Thompson, S. J. and Johnstone, C. J. and Thurlow, M. L. (June 2002). Universal Design Applied to Large Scale Assessments
- Willi Kunz (1998), Typography: Macro- + Microæsthetics, Verlag Niggli AG [1998] ISBN: 3-7212-0348-8

6. Annexure

-

6.1 Test Version 1

Name Contect Age Contect Gender Enal • Presse solve all 10 questions Time : 30min	5. What number should replace the question mark?	First 1: All dogs like to run. First 2: Some dogs like to run. First 2: Some dogs like to run. First 2: Some dogs like to run matters. If the first three statements are fact, which of the following statements must also be a fact? I: All dogs who like to waits look like their matters. I: Dogs who like to waits look like their matters.
Our is to revelous as foot is to : a. Running Seaster Satestee d; Starteboard Jumping	6. Tanya is odder than Teric. Cliffs older than Tanya. Eric is older than Cliff. If the first two statements are true, the third statement is	III: Dogo who like to run do not kook like their mesters. a. I only b. II only c. II and III only d. None of the statements is a known fact
2. Choose which of the five words is the odd one out. a) Satisfactory b) Perfect c) Acceptable d) From	a) True a) False b) Uncertain	
e) sources 3. What letters should appear in the fourth square? $ \begin{array}{c c} \hline A & C \\ \hline G & E \end{array} & \hline K & H \end{array} $	7. GEAN: MERG::STAR::	
4. Confound is to bewilder as astround is to: a) Securitie 4) Confuse 4) Securitie 4) Astoritie 4) Astoritie 4) Henrity	Which is the odd one out? a) Congregation b) Dispersion Consultation d) Consolution d) Consolution ei Aggregation	
Page 1	Page2	Page 1

6.2 Test Version 2

Name Contact Contact Gender Email	Time : 30min	TANYA is adder than ERIC CLIFF is adder than TANYA TOUC - or on CLIFF	7	යිඩම is to MBG as මී ලිමුම් is to	a. TSRA b. RTSA c. RTAS d. TARS
1 CONFO2AD is to DEDULIN	a. Surprise b. Confuse c. Startle d. Astonish e. Horrify	EHL is dider than CCTTT If the first two statements are true, the third statement is a. True b. False c. Uncertain	8	which is the $_{\rm odd}$ one out?	a. Congregation b. Dispersion c. Compilation d. Convocation e. Aggregation
Choose which of the words is the odd one	five a. Satisfactory out b. Perfect c. Acceptable d. Fine e. Suitable	(e) all Dags like to RUN 20000 Dags like to RUN		What Number should Replace	
(3) What letters sho FOURTH SCO AC BE KH OK	uld appear in the NET 7 77	If the first three statements are facts, which of the following statements must also be a fact? I. All dogs who like to swim look like their masters. II. Dogs who like to swim also like to run. III. Dogs who like to run de not look like their masters.		the Question 5 3 mark? 7 3 2	
(4) Oar is to Rowboar FOOT is to	a. Running b. Sneaker c. Skateboard d. Jumping	a. I only b. II only c. II and III only d. None of the statements is a known fact.			

6.3 Test Version 3

