An Empirical Study of the Effects of Personality on Software Testing

Tanjila Kanij and John Grundy

Swinburne University of Technoogy

Melbourne, Australia

Robert Merkel

Monash University

Melbourne, Australia

Outline

- Motivation

- Background

- Experimental design

    - Personality Assessment
    - Testing Task and performance assessment

- Results

- Conclusions

Motivation

- Personality impact has been studied on various aspects of Software Engineering e.g. coding, pair programming, teamwork

- Anecdotally, it has been thought that software testers are more conscientious, neurotic, more open to experience…

- But no one really knows!

- In an earlier study of professional testers opinions, we found mixed views on what impacts testing performance and ability

Background

- Personality – MBTI vs Five Factor model

- Human factor impact on software testing
  - Experience, attitude, organisational impact

- Personality factors and programming
  - Specific MBTI traits over-represented, but…
  - Five factor-based assessment suggested no predictors

- Personality factors and software engineering
  - Capretz's studies – sensing, thinking, judging, intuitive critical factors (MBTI); Feldt et al's "clusters" of factors
  - Clark et al's conscientiousness, introversion findings

"Expert Opinion"

- Armour – "nose for testing"

- Pettichord – tolerate tedium, skeptical, handle conflicts

- Pol et al – creative, accurate, strict in methodology

- Capretz and Ahmed – job responsibility analysis – attention to detail, good organisational skills, sensing and judging

Our Study

- Empirically determine relationship between personality type using Five Factor model and testing performance

- Use Computer Science & Software Engineering students as the population to sample

- Quasi-experiment of:

  - testing task to complete
  - personality assessment
  - performance assessment

Assessment of Personality

- Five Factor model

- NEO PI-3 inventory, measuring:

  - **Extraversion (E):** related to sociability, assertiveness, talkativeness and activeness.

  - **Agreeableness (A):** the expressive quality of admirable human aspects of personality

  - **Conscientiousness(C):** "Will to achieve" - purposeful, strong-willed and determined

  - **Neuroticism (N):** covers forms of excessive emotionality. Facets of this include anxiety, angry hostility, depression, self-consciousness, impulsiveness and vulnerability.

  - **Openness to Experience (O):** Openness to Experience is associated with intelligence and intellectual interests.

Testing Task, Performance Assessment

- Test faulty Java program (derived from assignment in another unit)

- 18 de-identified assignments used to craft one with common (and uncommon) faults – 20 in all; 1017 lines code; max method cyclometric complexity of 7

- Classified severity using Hutchison's taxonomy

- Compared injected bugs to Knuth's errors and Eisenstadt's bug war stories

Faults & Classification

Bug No.	Correct program behaviour	Incorrect program behaviour (Bug)	Fault description	Fault Source	Severity	Knuth's errors	Eisenstadt's bug war stories
1	The output is printed in the console and in the file	The output is printed only on the console	Missing functionality	Students	Low	forgotten function	unsolved
2	Pair of cities are printed with their names	Pairs are printed with numbers	Wrong implementation	Students	Medium	algorithm awry	
3	Prints right number of pairs possible among the given cities	Prints wrong number of pair of cities	Wrong implementation	Researchers	High	blunder or botch	
4	Prints right distance between the pair of cities	Prints wrong distance between pair of cities	Wrong function call	Researchers	High	mismatch between modules	des.logic
5	Prints pairs with right (name/number) of cities	Prints pairs with wrong (name/number) of cities		Researchers	High	blunder or botch	
6	Prints right number of topologies possible among the given cities	Prints wrong number of topologies among the given cities	Wrong index	Students	High	blunder or botch	lex/var
7	Prints right topology distance	Prints wrong topology distance	Wrong index	Students	High	blunder or botch	lex/var
8	Selection of right shortest distance	Selection of wrong shortest distance	Wrong parameter	Students	High	blunder or botch	var
9	Not printing duplicate topologies	Printing duplicate topology		Researchers	Medium	blunder or botch	
10	Right display of distance unit	Wrong display of distance unit	Wrong string literal	Researchers	Medium	trivial typo	
11	Should calculate distance for all values	Cannot calculate topologies for special value (gives exception with 2 cities)		Students	Critical	Forgotten function	lang
12	Calculates right distance for cities with <0 degree	Calculates wrong distance for cities with <0 degree	Missing Bracket	Students	High	trivial typo	lex
13	Should work with any number of cities	For more than 6 cities gives memory limit exceed exception	Memory	Students	High	Language liability	mem

Assessment Metrics

- Bug location rate (BLR):
  - number bugs found / time taken (mins)

- Weighted fault density (WFD):
  - sum of (weight * severity) / number found

- Bug report quality (BRQ):
  - assessed using the IEEE standard of Test Documentation

- Overall effectiveness
  - Total score (BLR+WFD+BRQ) vs
  - Weighted total score (0.3*BLR+0.3*WFD+0.4*BRQ)

Results

- 48 students; 18-35 years old; 69% male

- 23% had professional experience in testing

- 31% had done specialised testing unit

- 27% had used testing tools

- Shapiro-Wilk Test indicated that our population distributions do not differ significantly from normality, for the NEO personality inventory used to assess personality

Distribution of Scores

Table 2: Distribution of scores (N = 48)

	Minimum	Maximum	Average	Std
Neuroticism (N)	32	74	54.94	10.38
Extraversion (E)	20	70	50.42	9.32
Openness to experience (O)	35	80	54.36	10.02
Agreeableness (A)	29	74	48.27	9.62
Conscientiousness (C)	27	66	47.25	8.62
O_{sum}	0.63	3.87	1.91	0.85
O_{wsum}	0.24	1.51	0.73	0.34
Bug Location Rate (BLR)	0.02	0.37	0.12	0.063
Weighted Fault Density (WFD)	0.1	0.33	0.23	0.07
Bug Report Quality (BRQ)	0.5	3.5	1.56	0.83

Personality traits vs testing effectiveness

Table 3: Correlations (N = 48)

	N	E	O	A	C	O_{sum}	O_{wsum}	BLR	WFD	BRQ
Neuroticism	1	-0.329	-0.136	-0.135	-0.457	0.034	0.036	0.122	0.02	0.043
Extraversion		1	0.401	-0.231	0.375	-0.267	-0.267	0.038	-0.133	-0.0191
Openness			1	-0.235	0.177	0.161	0.165	-0.025	-0.154	0.179
Agreeableness				1	-0.121	0.167	0.173	-0.034	-0.215	0.191
Conscientiousness					1	0.026	0.026	0.251	-0.241	0.028
O_{sum}						1	1.000**-	0.258	0.085	0.996
O_{wsum}							1	0.248	0.071	0.998
BLR								1	-0.310	0.214
WFD									1	0.028
BRQ										1

Outcomes

- Weak negative correlation – extraversion vs overall effectiveness (differs from previous studies)

- Weak negative correlation – extraversion and bug report quality - surprising?

- Weak positive correlation – conscientiousness and bug location rate – expected?

- Weak negative correlation - conscientiousness and weighted fault density – more vs severity (quantity vs quality?!)

Implications

- Who makes a better tester – does personality matter???

- Need to be conscientious ☺

- Extroversion-related qualities might negatively impact bug reporting?!

- Teaching testing – bug location vs bug severity vs bug report quality

  - Not all bugs are equal!

- Assessing testing – when know student / tester has done a good job??

Summary

- Empirical study of CS&SE students to examine impact of personality, as measured by Five Factor model, on testing effectiveness

- Moderate size Java program with 20 errors, ranging in severity, derived from older student exemplars & widely used standard

- Most personality indicators didn't seem to impact testing effectiveness in our study

- Weak +ve impact of conscientiousness on finding bugs, but –ve on severity – quantity vs quality??

- Weak –ve impact of extraversion on effectiveness

References

Kanij, T., Merkel, R., and Grundy, J.C., An Empirical Study of the Effects of Personality on Software Testing, 2013 International Conference on Software Engineering – Education and Training (CSEET2013), San Francisco, USA, May 19-21, 2013, IEEE CS Press.

Kanij, T., Merkel, R., Grundy, J.C. Some lessons learned from conducting industry surveys in software testing, ICSE2013 Workshop on Conducting Empirical Studies in Industry, San Francisco, USA, 20th May 2013.

Kaniji, T., Merkel, R. and Grundy, J.C. Assessing the Performance of Software Testers, in 3rd ICSE workshop on Collaborative and Human Aspects of Software Engineering, Zurich, Switzerland, 2nd June 2012.

Kanij, T., Merkel, R. and Grundy, J.C. A preliminary study on factors affecting software testing team performance, In Proceedings of 2011 International Conference on Empirical Software Engineering and Methods (ESEM 2011), Sept 19-23, Banff, Canada , IEEE Press.