

Abwägung bei der Methodenwahl: PP und TFD verbessern nicht alles Beiträge zum Softwareengineering

Motivation und Spezifikation

•an Universität bekommt man Gefühl: Hauptsache agil

•Frage: Bewirken einzelne Methoden tatsächlich nur Verbesserungen?

•Frage: Wie wirken sich Methoden auf Tests und Korrektheit aus?



Quelle: www.googlefight.com

Fahrplan

- notwendige Begriffe
- Betrachtung von drei Veröffentlichungen

1. Einfluss von PP auf Testqualität:

„On the Effects of Pair Programming on Thoroughness and Fault-Finding Effectiveness of Unit Tests“, L. Madeyski, 2007

2. Einfluss von TFD auf Testqualität:

„The impact of Test-First programming on branch coverage and mutation score indicator of unit tests:An experiment“, L. Madeyski, 2009

3. Einfluss von PP und TFD auf Hilfsfunktionen:

„Development of Auxiliary Functions: Should You Be Agile? An Empirical Assessment of Pair Programming and Test-First Programming“, O. Lemos et al., 2012

- Zusammenfassung

- Diskussion

Definitionen (I)

- Pair-Programming
 - jede Entwicklungstätigkeit als Team durchzuführen
- Test-First-Development (auch Test-Driven-Development)
 - vor jeder Implementierung muss ein Test dafür geschrieben werden

Definitionen (II)

- Branch Coverage
 - stärker als Anweisungsüberdeckung
 - prüft, ob an Verzweigungen alle Folgen berücksichtigt werden
- Mutation Score Indicator
 - erzeuge automatisiert Mutationen des Codes
 - prüfe Anzahl Mutanten, die durch Tests erkannt werden
 - zeigt, ob Tests für den Code ausreichen

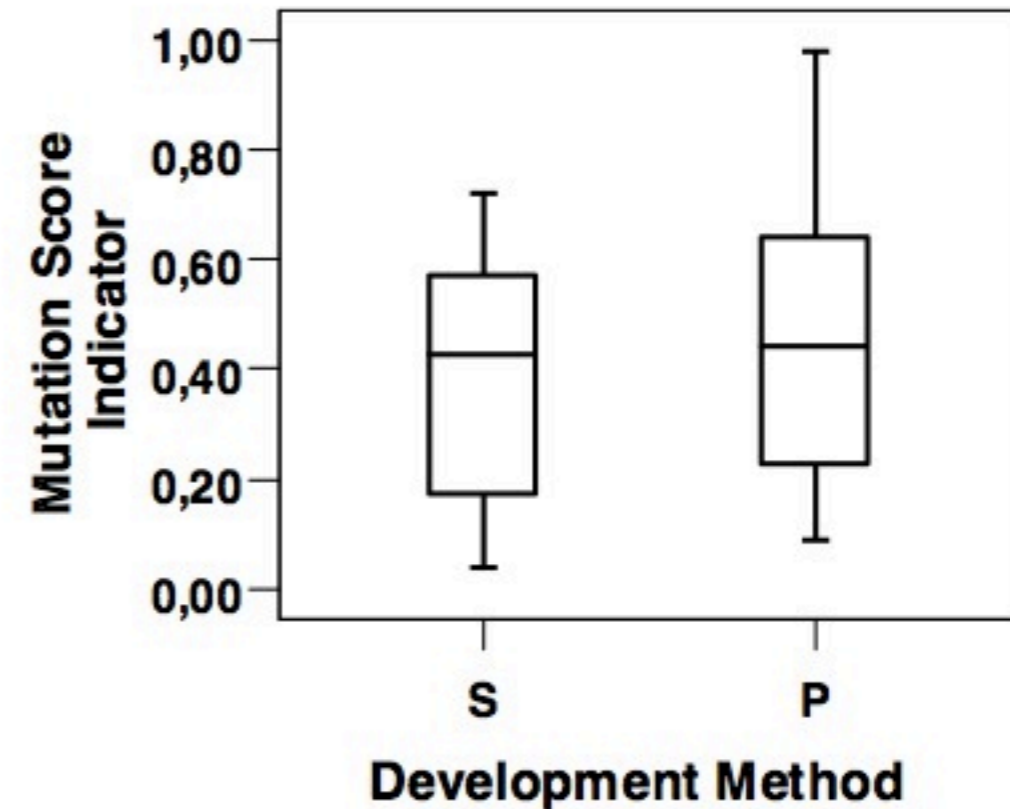
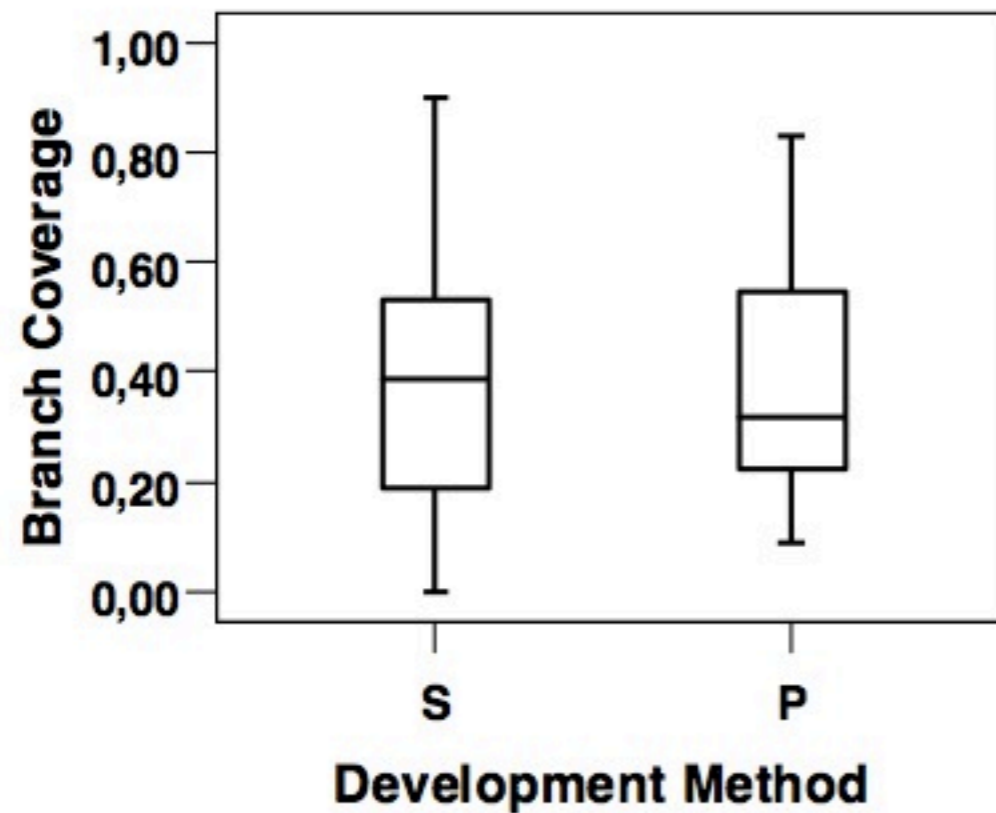
Definitionen (III)

- Functional Test Set Success Level (FTSSL)
 - I: kein Test erfolgreich
 - N: nicht alle Tests erfolgreich
 - C: alle Tests erfolgreich

Einfluss von PP, Experiment

- Informatikstudenten, zweites und drittes Jahr
- 28 Solo-, 35 Pair-Programmer
- User Stories als Spezifikation (Abrechnungssystem)
- Java, Eclipse, JUnit, TFD
- Vorbereitung
 - sieben Vorlesungen
 - sieben Labortermine (à 90 min.)
- messe Branch Coverage und Mutation Score Indicator

Einfluss von PP, Ergebnisse



Quelle: (1)

- kein signifikanter Unterschied in BC und MSI
- PP hat keinen Einfluss auf Abdeckung

Einfluss von PP, Validität und Kritik

• positiv:

- große Anzahl Testpersonen
- wiederholbar
- einfache, wirkungsvolle Metriken

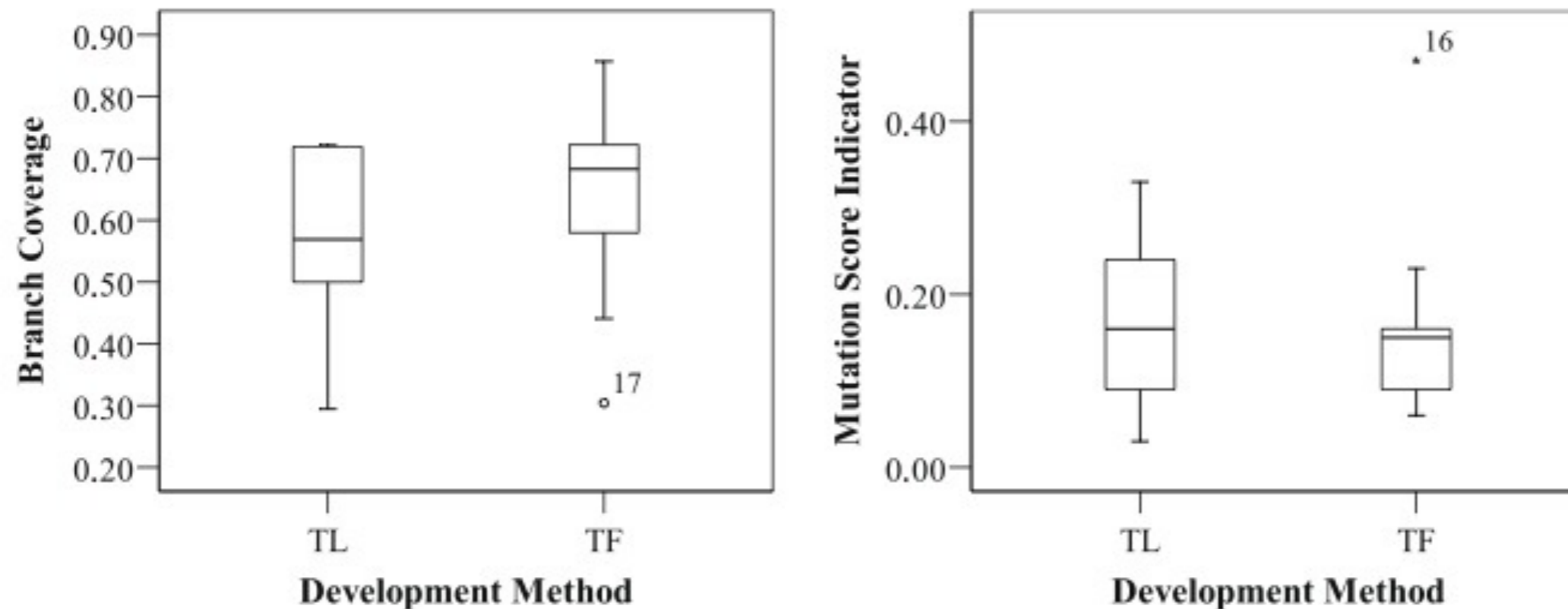
• negativ:

- Studenten (ohne Abschluss)
- erreichte BC und MSI grundsätzlich sehr niedrig
- nur eine Implementierung vorgenommen

Einfluss von TFD, Experiment

- 22 Informatikstudenten (Master), >450 h Programmiererfahrung
- User Stories als Spezifikation (webbasiertes Abgabesystem)
- Java, Eclipse, JUnit
- Vorbereitung
 - Vorlesungen in Web-Development
 - drei zweistündige Trainings in TFD und TLD
- Vorerfahrung bei Gruppeneinteilung berücksichtigt (Fragebogen)
- messe Branch Coverage und Mutation Score Indicator

Einfluss von TFD, Ergebnisse



Quelle: (2)

- 9 Studenten TFD (3 ausgeschlossen), 10 Studenten TLD
- zu viele Freiheitsgrade, Ergebnisse nicht signifikant

Einfluss von TFD, Kritik

- positiv:

- Gruppen durch Fragebogen sehr homogen
- qualifizierte Studenten

- negativ:

- zu geringe Anzahl Testpersonen
- TFD wurde für schwieriger empfunden
- Projektzeit zu kurz
- nur eine Implementierung
- Menge an Freiheitsgraden zu groß

Einfluss von PP und TFD auf Hilfsfunktionen, Experiment

- Informatikstudenten
 - für PP 46 Zweitsemester, 50-minütiges Training
 - für TFD 39 Sechstsemester, 2x 100-minütiges Training
 - keine Kombination beider Methoden
- schreibe einfache Hilfsfunktionen (10 - 200 loc), nicht in Java API
 - mathematische Operationen, Array- und Stringmodifikation
- Referenzgruppe mit 7 „Professionals“
- messe Coverage, Anzahl an Tests und FTSSL

Einfluss von PP und TFD auf Hilfsfunktionen, Ergebnisse (I)

		PP			Freq.
		I	N	C	
SP	I	0	2	1	3 (6.5%)
	N	0	33	7	40 (87%)
	C	0	3	0	3 (6.5%)
Freq.		0 (0%)	38 (82%)	8 (18%)	

		TF			Freq.
		I	N	C	
TL	I	0	2	0	2 (5%)
	N	2	28	2	32 (82%)
	C	0	3	2	5 (13%)
Freq.		2 (5%)	33 (87%)	4 (7%)	

Quelle: (3)

- mit PP mehr korrekte Lösungen als mit SP
- mit TFD kein signifikanter Unterschied in Verteilungen

Einfluss von PP und TFD auf Hilfsfunktionen, Ergebnisse (II)

Subj.	FTSSL		TDT		Cov		# TC		Subj.	FTSSL		TDT		Cov		# TC		Subj.	FTSSL		TDT		Cov		# TC	
	TL	TF	TL	TF	TL	TF	TL	TF		TL	TF	TL	TF	TL	TF	TL	TF		TL	TF	TL	TF	TL	TF	TL	TF
1	N	N	35	45	0%	100%	0	8	14	N	N	24	24	100%	100%	5	5	27	N	N	40	20	100%	100%	5	3
2	N	N	37	44	0%	100%	0	3	15	N	I	16	60	0%	75%	0	1	28	N	N	25	30	100%	100%	2	3
3	N	N	22	33	0%	75%	0	3	16	N	N	10	15	0%	100%	0	5	29	N	C	8	25	100%	89%	4	4
4	N	N	15	35	0%	100%	0	4	17	N	N	46	60	0%	81%	0	5	30	C	N	26	48	0%	91%	0	1
5	N	N	13	25	0%	100%	0	4	18	N	N	50	23	0%	100%	0	7	31	N	N	20	37	100%	100%	4	5
6	N	N	20	21	100%	88%	3	4	19	N	N	47	21	100%	100%	2	3	32	N	N	39	12	100%	100%	4	7
7	N	N	18	16	100%	100%	6	2	20	N	N	20	27	0%	100%	0	1	33	N	N	20	32	70%	88%	4	4
8	N	N	15	14	100%	100%	5	5	21	C	N	18	37	88%	100%	2	6	34	N	N	11	12	89%	100%	5	4
9	N	N	12	15	100%	100%	2	3	22	N	N	11	19	0%	100%	0	3	35	N	I	32	40	0%	100%	0	1
10	N	C	33	33	100%	100%	4	8	23	N	N	8	23	0%	100%	0	4	36	N	N	17	33	100%	100%	7	5
11	C	C	29	25	100%	100%	5	5	24	N	N	23	27	100%	100%	4	5	37	N	N	38	33	100%	100%	1	5
12	N	N	45	15	100%	100%	0	7	25	C	C	18	21	100%	100%	6	14	38	I	N	55	35	100%	100%	1	5
13	I	N	15	10	84%	100%	2	3	26	N	N	40	40	0%	100%	0	4	39	C	N	30	50	0%	83%	0	7
Mean (considering all subjects)																			25.67	29.10	57%	97%	2.13	4.51		

Quelle: (3)

- mit TFD mehr Testfälle und höhere Coverage
- mit TLD oft keine Testfälle

Einfluss von PP und TFD auf Hilfsfunktionen, Ergebnisse (III)

Subj.	TDT		FTSSL		Cov		# TC	
	TL	TF	TL	TF	TL	TF	TL	TF
01	31	48	N	N	100%	100%	1	1
02	13	35	N	N	86%	100%	1	1
03	50	50	I	N	0%	0%	0	1
04	31	35	N	N	0%	0%	0	0
05	23	28	N	N	100%	100%	2	3
06	37	23	N	N	100%	100%	3	5
07	50	35	N	N	100%	100%	7	6
Mean	33.57	36.28	–	–	69%	71%	2	2.43

Quelle: (3)

- „Professionals“ haben keine korrekte Lösung erzielt
- Anzahl der Testfälle sehr gering
- Coverage z.T. 0%

Einfluss von PP und TFD auf Hilfsfunktionen, Kritik (I)

- positiv:

- große Teilnehmerzahl
- verschiedene Aufgaben
- jeder Student implementiert mit beiden Methoden

Einfluss von PP und TFD auf Hilfsfunktionen, Kritik (II)

- negativ:

- möglicherweise zu leichte Funktionen (Unterforderung)
- Studenten (ohne Abschluss)
- sehr kurzes Training
- Testen offenbar nicht verstanden (0% Cov, geringe Testanzahl)
- „Professionals“?
- Coverage anfechtbar, FTSSL zu strikt?

Zusammenfassung PP

- kein signifikanter Einfluss auf BC und MSI
 - Qualität der Tests bleibt gleich

- mehr als doppelt so viele korrekte Implementierungen
 - im Experiment dennoch sehr wenig

Zusammenfassung TFD

- TFD positiv für Anzahl Tests und Coverage
 - MSI leider unbekannt

- Korrektheit der Implementierung nicht beeinflusst

Abschluss

- PP und TFD lösen nicht alle Probleme
- nicht immer kompletter Umstieg auf agil möglich
 - für Projekt prüfen, welche Methode nötig ist
- Kombination der Methoden in diesen Papern nicht berücksichtigt

Fragen/Feedback

- Diskussionseinstieg: Sehen Sie/seht ihr agile Methoden in der Summe als Nonplusultra?

Quellen

- (1) „On the Effects of Pair Programming on Thoroughness and Fault-Finding Effectiveness of Unit Tests“, L. Madeyski, 2007
- (2) „The impact of Test-First programming on branch coverage and mutation score indicator of unit tests: An experiment“, L. Madeyski, 2009
- (3) „Development of Auxiliary Functions: Should You Be Agile? An Empirical Assessment of Pair Programming and Test-First Programming“, O. Lemos et al., 2012